This book is presented

by

The Government of the United States

as an expression of

Friendship and Goodwill

of the

People of the United States

towards

The People of India

AN ADVANCED COURSE IN
GENERAL COLLEGE PHYSICS

·The M Co·

THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO · DALLAS
ATLANTA · SAN FRANCISCO

MACMILLAN AND CO., LIMITED
LONDON · BOMBAY · CALCUTTA · MADRAS
MELBOURNE

THE MACMILLAN COMPANY
OF CANADA, LIMITED
TORONTO

AN ADVANCED COURSE IN GENERAL COLLEGE PHYSICS

BY

PAUL LEVERNE BAYLEY, PH.D.

Professor of Physics
Lehigh University

AND

CHARLES CLARENCE BIDWELL, PH.D.

Professor of Physics
Lehigh University

NEW YORK

THE MACMILLAN COMPANY

Published September, 1936.
Fourth printing January, 1948.

SET UP AND ELECTROTYPED BY J. S. CUSHING CO.
PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

This text has been prepared primarily for college students who have had a previous course in Physics. As soon as a student has had a descriptive course in Physics he is then prepared either to widen his knowledge at the same level or to delve deeper into those subjects with which he has become somewhat familiar. To students without mathematical training only the former field of study is open. For them many textbooks have been written. Most of those texts use only trigonometry although a few give bare outlines of a calculus proof here and there. However, for the student who has had analytic geometry, who knows the meaning of a derivative, and is ready to examine more deeply into the nature of things, there seem to be no textbooks except in special fields. The authors have prepared this text in General Physics for those students.

Whenever material from a first course in Physics is required in building up a logical framework, that material has been included in this text, but a definite attempt has been made to include no more of it than is essential.

It is hoped that the text will be valuable to students majoring in Physics and for engineering students who alike must get a thorough training in the fundamentals of Physics before proceeding further. It is expected that all such students will be taking simultaneously a course in calculus, and that they have had the elements of the process of differentiation. They should understand that a derivative represents a "rate of change" and know the derivative of x^n. The calculus is introduced into the text gradually so that the derivatives of logarithmic and trigonometrical functions are not used until after they are reached in a calculus course. Integration is introduced at once as the inverse of a differentiation, the definite integral being obtained by evaluating the integration constant. Then in special simple cases where a summation is necessary it is shown that the results may be obtained by substituting limits into the results of an integration. Thus the definite integral is set up, shown to represent a summation, and

used long before the more rigorous proof can be reached in the calculus course.

Because of the necessary duplication of many elements of a first course in Physics, the text could be used as a first course for students who have had a good background of mathematics.

The problems have been selected so as to give practice in both metric and English units. They are largely numerical but a few are theoretical. In every list a few are simple and a few are difficult enough to challenge the better students. It is intended that the problems will be fairly difficult for the average student.

P. L. BAYLEY
C. C. BIDWELL

BETHLEHEM, PA.
August, 1936

CONTENTS

CHAPTER I. MECHANICS

CHAPTER III. HEAT

CHAPTER IV. ELECTRICITY AND MAGNETISM

AN ADVANCED COURSE IN GENERAL COLLEGE PHYSICS

CHAPTER I

MECHANICS

Fundamental Concepts of Position and Motion

1. Displacement and Vector Quantities. — The position of a particle with reference to some arbitrary origin is usually designated either by means of rectangular co-ordinates, x, y, and z, or by means of vectors. The line drawn from the origin to the particle is called the position vector, s. When the position of a particle is changed, the displacement is given by specifying the changes Δx, Δy, Δz, or by Δs, the change in the position vector. All physical quantities may be classified as either scalar quantities or vector quantities, the former having magnitude only, the latter magnitude, direction, and sense. A vector quantity may be represented by a line whose length and direction indicate respectively the magnitude and direction of the quantity. An arrowhead is always placed on the end of the line to show the "sense" of the direction. We may have position vectors such as OA and OC (Fig. 1), locating the positions of points A and C with reference to some origin O. When a displacement occurs from O to A, we call the displacement vector the vector drawn from the point representing the initial position of a particle to the point representing the final position. We have velocity vectors, lines specifying the direction and magnitude of velocities, acceleration vectors, force vectors, and various others.

2. Vector Operations. — To add two vectors, a and b, lay off the vector a from some point O (Fig. 1) and then lay off the vector b starting from the end of vector a. The line from O to C is

defined to be the sum of the two vectors. The vector difference $a - b$ is considered to be the vector sum of a and $-b$ as in Fig. 1. If a represents one displacement from O and b another displacement, then the vector from O to C represents the resultant displacement and likewise for acceleration vectors, etc. The resultant displacement may also be computed by trigonometry, since the vectors form the two sides of a triangle with lengths and directions known.

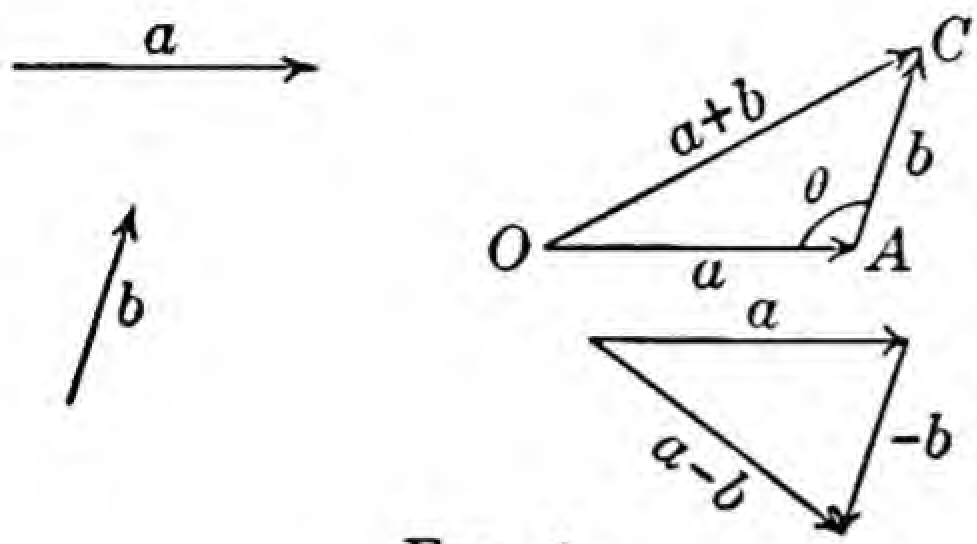


FIG. 1

In Fig. 1, $OC^2 = a^2 + b^2 - 2\,ab\cos\theta$. If several like vectors, when placed end to end, form a closed polygon, the resultant vector, that is, the sum of all the vectors, is zero (Fig. 2). This is a test which may be applied in the case of balanced forces; the sum of all the forces vectorially added must give a closed figure. This is discussed in § 14.

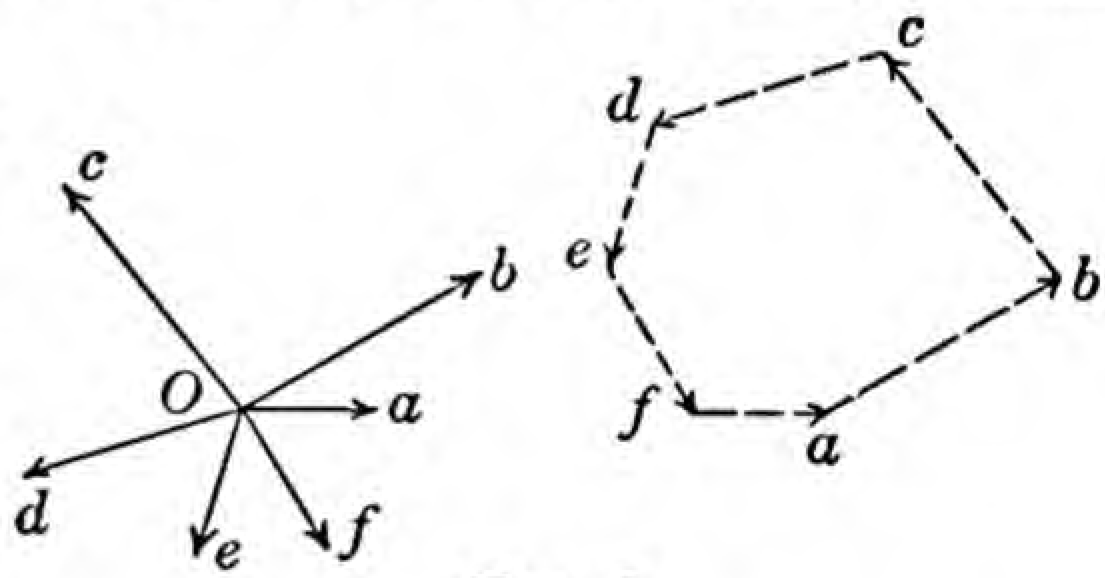


FIG. 2

A single vector may be resolved into any number of component vectors, the only condition being that the sum of the components, vectorially added, must be equal to the original vector. The vector OC (Fig. 1) may be resolved into vectors OA and AC, and is equivalent to these two. We often are concerned with the multiplication of two physical quantities which may be represented by vectors. In fact, we find two separate cases. Consider the physical quantity work, which is a scalar. It is computed either by the product of the vector force F and the component of the displacement vector which is in the direction of the force or by the product of that component of the vector force which is in the direction of the displacement and the displacement s, that is, by $Fs\cos\theta$, where θ is the angle between the two vectors. The vector $s\cos\theta$ is parallel to F. So we see that the product of two parallel vectors or the product of two non-parallel vectors and the cosine of the angle between them gives a scalar quantity.

Now consider the moment of a force. Let a force F (Fig. 3) be applied at A to a body. Let O be some point within the body and let s be the position vector of A with respect to O. The tendency

which the force F has in producing rotation about any axis through O depends on the magnitude of F and upon the perpendicular distance l from O to the line of action of the force. Therefore it has been agreed to measure the moment of a force about a point by the product of the force and the perpendicular distance l from the point to the line of action of the force. This moment could produce no rotation about an axis through O in the plane of s and F and would produce a maximum angular acceleration about an axis through O perpendicular to that plane. Therefore the moment of force is seen to have a certain directive property as well as a magnitude and may be represented by a vector. The magnitude of the vector is $L = Fs \sin \theta$ and the direction is chosen as that of the perpendicular to the plane determined by F and s. The sense of the direction is arbitrarily chosen to be that in which a right-handed screw would be advanced under the action of the given moment of force.

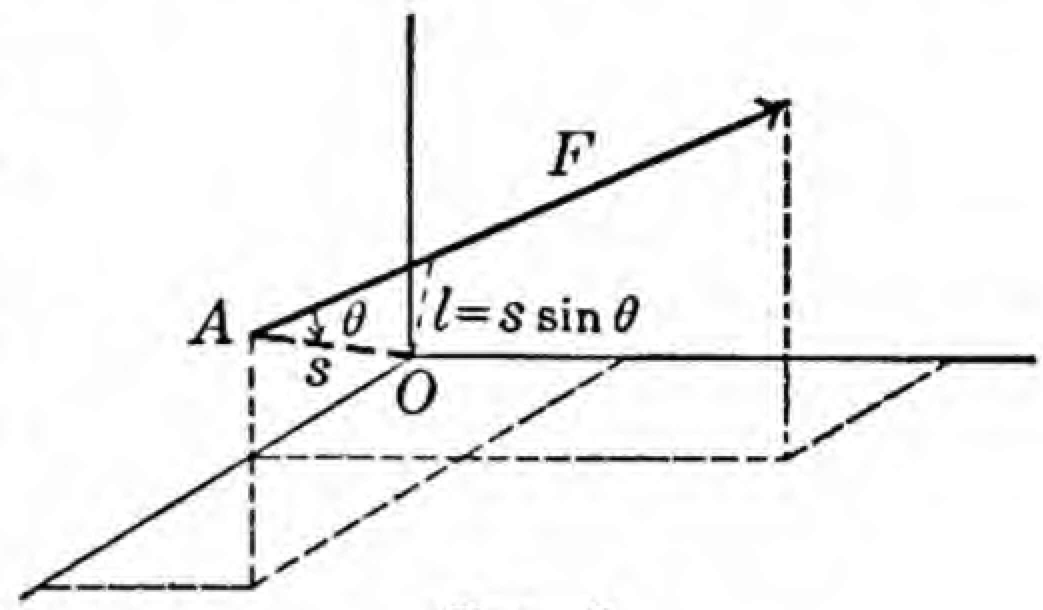


FIG. 3

3. Velocity and Acceleration. — If s represents the displacement of a particle from some origin (in such a case the displacement and position vectors are the same), then ds/dt, the rate of change of position with time at some particular instant, is called the velocity v at that instant. *Both magnitude and direction must be expressed in specifying a velocity.* We call the *speed* of a body the magnitude of its velocity without reference to its direction.

If v represents the velocity of a particle, then dv/dt, the rate of change of velocity with time at a particular instant, is called the acceleration a at that instant. Magnitude and direction must be specified for the acceleration also. In mathematical language we may express these definitions as

$$v = \frac{ds}{dt}, \tag{1}$$

$$a = \frac{dv}{dt} = \frac{d^2s}{dt^2}. \tag{2}$$

From Eq. 2 we see that since the acceleration is the derivative of v taken with respect to the time t, then in order to find the velocity

v, we must reverse the process of differentiation, *i.e.* take the integral of the acceleration a with respect to the time t. From Eq. 2,

$$dv = a\,dt,$$

and
$$v = \int a\,dt. \tag{3}$$

Likewise, if v is given, we have the function which is the time derivative of the displacement s. So s may be obtained by integration. From Eq. 1,

$$ds = v\,dt,$$

and
$$s = \int v\,dt. \tag{4}$$

Eqs. 3 and 4 are general equations fitting any type of motion. Integration, however, may not be performed until v and a are expressed in terms of t. In other words we must know how the velocity or the acceleration varies with the time.

4. Uniformly Accelerated Motion in a Straight Line. — Motion in which the acceleration is constant in magnitude and direction is called uniformly accelerated motion. Because a is constant, integration of Eq. 3 gives

$$v = at + c,$$

where c is the constant of integration. If we consider the process of integration as just the inverse of differentiation, we attempt to find the function which, when differentiated with respect to the time, will give the acceleration a. One solution is evidently $v = at$, but we shall see that this is only a special solution for the motion of a body starting from rest. The physical meaning of c is found by introducing what are called *the initial conditions*. Suppose that the body undergoing uniformly accelerated motion was moving with a velocity of v_0 when observations are started, at a time called $t = 0$. The initial conditions are: when $t = 0$, $v = v_0$. Substituting these in the above equation, we obtain $c = v_0$. Therefore, we have

$$v = at + v_0. \tag{5}$$

If this value of v be substituted in Eq. 4, we have

$$s = a\int t\,dt + v_0\int dt. \tag{6}$$

By integration we obtain

$$s = \tfrac{1}{2}\,at^2 + v_0t + c', \tag{7}$$

where c' is the constant of integration. In this equation it is necessary to give a value of s when $t = 0$, in order to evaluate c'. If we choose to measure the distance traveled by the body from its position when we start counting time, then we have $s = 0$ when $t = 0$. Under these conditions, we have $c' = 0$ and

$$s = \tfrac{1}{2}\,at^2 + v_0t. \tag{8}$$

Eliminating t from Eqs. 5 and 8,

$$v^2 = v_0^2 + 2\,as. \tag{9}$$

The student should eliminate a from the same two equations and from the resulting equation see the use which can be made of the average velocity $\dfrac{v + v_0}{2}$.

PROBLEMS

1. Following the method employed in the case of uniformly accelerated motion, discuss the case where $a = 0$ (uniform motion), finding v and s.

2. Find the equations for and draw curves representing (1) velocity, and (2) distance covered from starting point, as a function of time, for the following cases, in each case considering that the body was traveling 10 cm./sec. when observations were started and that the body had traveled 5 cm. before that time:

(*a*) acceleration = 0.

(*b*) acceleration = 3 cm./sec^2.

(*c*) acceleration increasing directly with the time, such as $a = 4\,t$ cm./sec^2. In this last case, carry out the integrations as in Eqs. 3 and 4.

5. The Motion of a Projectile. — Suppose a body is projected into the air with an initial velocity v_0, at an angle θ with the horizontal. We shall find expressions for (*a*) the range (the horizontal distance traveled when the body reaches the ground); (*b*) the time of flight; (*c*) the maximum height; and (*d*) the equation of the path. It is convenient to resolve the motion into its horizontal and vertical components. The horizontal component is subjected to no acceleration (ignoring air friction); the vertical component has the regular acceleration due to gravity, $-g$, since the upward direction is taken as positive. We may then find the horizontal velocity v_x and displacement x by applying the expressions for

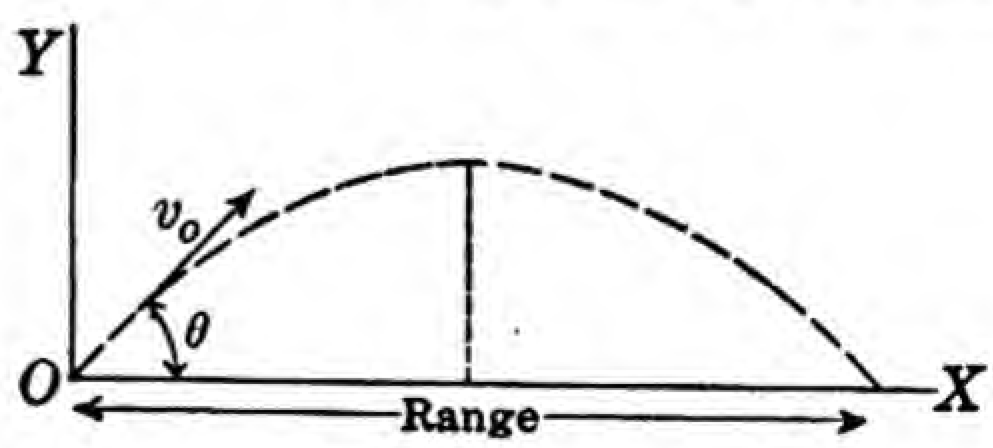


FIG. 4

uniform motion, and find the vertical velocity v_y and the displacement y by applying the expressions for uniformly accelerated motion. Thus, taking the origin at the starting point, we have for the horizontal component,

$$v_x = v_0 \cos \theta, \tag{10}$$

and

$$x = v_x t = v_0 \cos \theta \cdot t. \tag{11}$$

For the vertical component (see Eqs. 5 and 8),

$$v_y = (-g)t + v_0 \sin \theta, \tag{12}$$

and

$$y = \tfrac{1}{2}(-g)t^2 + (v_0 \sin \theta)t. \tag{13}$$

By substituting $y = 0$ in Eq. 13, we obtain two values for t. One of these values gives

$$\textit{the time of flight} = \frac{2\, v_0 \sin \theta}{g}.$$

If the value of the time of flight is substituted in Eq. 11, we obtain for the *range* the value

$$\frac{v_0^2 \sin 2\,\theta}{g}.$$

At the instant of maximum height the vertical velocity is zero. The time for the body to attain its maximum height may, therefore, be found by substituting $v_y = 0$ in Eq. 12. This value, as should have been expected, is half the time of flight (obtained above) and when substituted in Eq. 13 gives

$$\textit{the maximum height} = \frac{v_0^2 \sin^2 \theta}{2\,g}.$$

The *equation of the path* may be obtained by eliminating t between Eqs. 11 and 13:

$$y = x \tan \theta - \frac{g}{2\, v_0^2 \cos^2 \theta}\, x^2. \tag{14}$$

By transferring the origin of this curve to its highest point, that is, out on the x axis by the amount of half of the range and up the y axis by the maximum height, we get

$$y_1 = -\frac{g}{2\, v_0^2 \cos^2 \theta}\, x_1^2, \tag{15}$$

which is a standard form of the equation for a parabola.

For ordinary work, g may be taken as 980 cm./sec². or 32.2 ft./sec². (See §§ 11 and 67.)

6. Uniform Circular Motion. — Consider the motion of a particle traveling in a circle with constant speed. Let O be the center of the circle (Fig. 5). At a given instant, let the particle be in position (1) and after the lapse of a small time Δt, in position (2), having traveled along the arc a distance Δs. Let v_1 and v_2 represent the two velocities *equal in magnitude but differing in direction.* The velocity vectors may be placed in juxtaposition as shown, the vector Δv being $v_2 - v_1$. This vector has the direction indicated and its magnitude represents the change in velocity which takes place during the time Δt. It is to be carefully noticed that in uniform circular motion the velocity does not vary in the manner treated in § 4. In that case, the speed was variable and the direction was constant. In this case, the speed remains constant but the direction varies with the time.

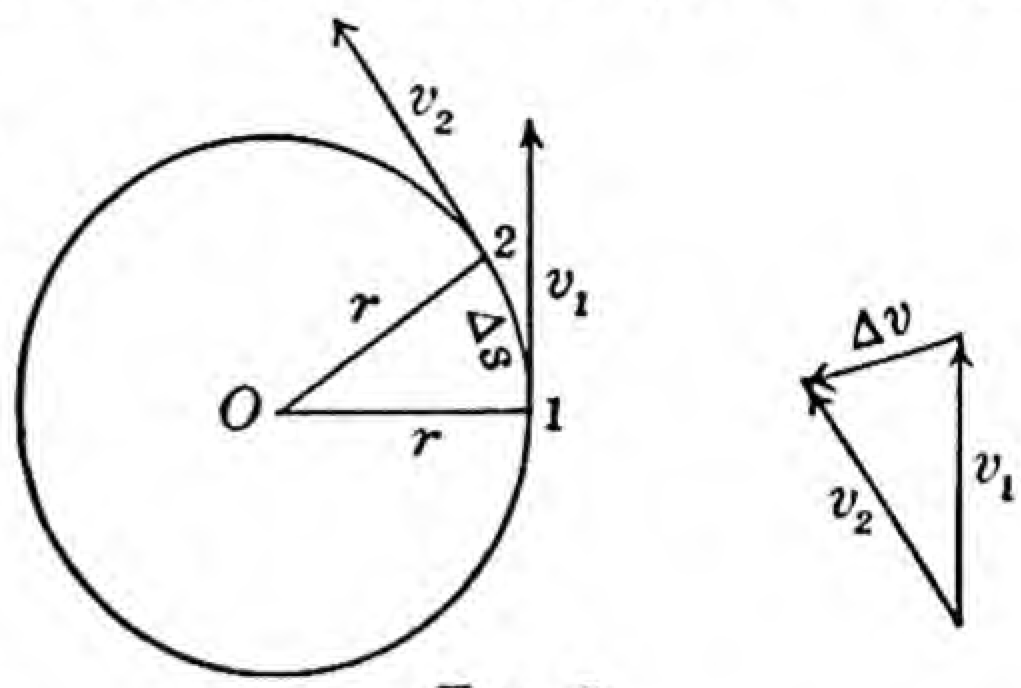


FIG. 5

It may be seen from the figure that as Δt increases the angle between v_1 and v_2 in the vector triangle increases and the angle which Δv makes with v_1 or v_2 changes. Now the average acceleration during the time Δt is $\frac{\Delta v}{\Delta t}$. Since Δt is a scalar quantity, it is obvious that $\frac{\Delta v}{\Delta t}$ is a vector having the same direction as Δv. We have just mentioned that Δv changes direction as Δt changes, so the expression

$$\frac{\Delta v}{\Delta t} = \frac{(v_2 - v_1)}{(t_2 - t_1)}$$

gives only an approximate value for the acceleration at some time t during the interval $(t_2 - t_1)$. The smaller the interval becomes, the more accurately does the expression give the instantaneous acceleration.

Since the points (1) and (2) must be taken very close together, the arc Δs may be considered as sensibly straight. The triangles having sides r, r, Δs and v_1, v_2, Δv respectively are similar (homologous sides being perpendicular). Hence

$$\frac{\Delta v}{v} = \frac{\Delta s}{r}. \tag{16}$$

Dividing each side by Δt we have

$$\frac{\Delta v}{v\Delta t} = \frac{\Delta s}{r\Delta t}. \tag{17}$$

Remembering that in the limit, as Δt approaches zero,

$$\lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} = a,$$

and

$$\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \frac{ds}{dt} = v,$$

we obtain from Eq. 17,

$$a = \frac{v^2}{r}. \tag{18}$$

The vector diagrams indicate that in the limit the direction of the acceleration is perpendicular to that of the velocity at any instant, the direction of the velocity being along the tangent, and the direction of the acceleration being toward the center. These qualitative facts may be seen from Newton's laws of motion (see next section). Since the motion is not in a straight line, some force must be acting. Since the velocity along the path is not increasing, no component of the force can possibly be parallel to the path. So the force must be at right angles to the path. The change in momentum must be directed the same as the force and hence the acceleration also. The acceleration as well as the force, therefore, is constant in magnitude although variable in direction.

PROBLEMS

1. A gun is mounted on a train at right angles to the direction in which it is moving with a velocity of 30 miles per hour. The muzzle velocity of a bullet shot from the gun is 500 ft./sec. At what angle with respect to the direction of motion of the train must the gun sight be set and at what angle of elevation must the gun be set in order to hit an object on the ground 5000 ft. from the train (measured along the ground under the projectile path)?

2. A train starts from rest and attains a speed of 60 miles per hour after traveling one mile. Find the average acceleration in miles per hour per second, miles per second per second, and feet per second per second.

3. A sphere 5 ft. in diameter has material sprayed on it which increases its diameter at the rate of one half an inch per hour. At what rate is its volume increasing?

4. A stone in a sling is whirled in a vertical circle. The cord of the sling is 75 cm. long. What is the minimum speed which will just prevent the stone from falling out of the sling?

5. In the case of a projectile show: (*a*) that the time of ascent and descent are the same; (*b*) that the initial and final vertical velocities are the same, but oppositely directed.

6. Show that, for any given value of v_0, the largest range is obtained for an angle of elevation of 45°. Show also that for any range short of the maximum there are two angles of elevation, one less than 45° and the other more than 45°, for which the projectile may be fired to reach a given point.

7. Obtain the equation of the path of a projectile fired horizontally from the origin with a velocity of v_0. Compare this equation with Eq. 15. Explain the difference.

8. A weight is dropped from an aeroplane traveling horizontally at a speed of 90 miles per hour. The aeroplane is at an altitude of 1000 ft. How long before the weight reaches the ground and what velocity does it acquire? What is the equation of the path and what sort of curve does this represent?

9. Bullets are fired from a mountain 5000 ft. higher than a valley below. The muzzle velocity is 500 ft./sec. Where will bullets hit in the valley for angles of fire of + 45°, 0°, and − 30°?

Newton's Laws of Motion

7. The First Law. — *If a body is at rest it will continue at rest or if in motion it will continue in motion with the same speed in the same straight line unless acted upon by some agency or influence external to the body.* This agency, which must act when any change in the motion of a body occurs, is called a force.

This law is based upon experimental evidence. It gives a qualitative definition of the term force, as the cause of any change in the motion of a body.

8. The Second Law. — From observation and experiment Newton stated that *the change in momentum of a body varies directly as the resultant force acting on the body and directly as the amount of time during which that force acts. The change in momentum always takes place in the direction of the resultant force.* These facts are stated in the following vector equation:

$$F\,dt = k\,d(mv), \text{ or } F = k\,\frac{d(mv)}{dt}. \qquad (19)$$

If the mass remains constant, we may write this equation in the more familiar (though less general) form:

$$F = kma. \qquad (20)$$

The value of k depends on the choice of units for expressing the magnitudes of F, m, and a.

9. Momentum and Impulse. — If we integrate Eq. 19, we obtain

$$\int_0^t F\,dt = k\int_{m_0v_0}^{mv} d(mv) = k(mv - m_0v_0). \qquad (21)$$

If F is variable, then the left-hand member cannot be integrated until F is expressed as a function of time.

The term $\int_0^t F\,dt$ is called the impulse of the force and so we might state the second law as follows: The change in momentum of a body is directly proportional to the impulse of the resultant force and takes place in the direction of that force.

Ordinary matter at ordinary velocities does not change in mass. However, at high speeds, approaching the velocity of light, we have experimental evidence of a large increase in mass. For ordinary velocities $m = m_0$ and the right hand member of Eq. 21 reduces to

$$\int_0^t F\,dt = km(v - v_0). \tag{22}$$

The first law of motion is obviously a special case of the second law. For the particular case where $F = 0$ then $mv = m_0v_0$. If the mass is constant, $m = m_0$ and $v = v_0$, so the body continues to move with constant speed in a straight line. Whether or not the mass of a body varies, we may state that when the resultant force on it is zero, the body continues to move with constant momentum in a straight line.

10. The Third Law. — Newton observed that no single force ever exists by itself. *Forces always occur in pairs.* They come into existence only through the mutual interaction of two bodies. If a body collides with a second body, simultaneously forces act on the two bodies pushing them apart. The forces increase to a maximum value when the two bodies are most distorted and decrease to zero when they separate, but at every instant during the collision the forces are equal to each other, and act in opposite directions on the separate bodies. (The student may find it interesting to study cases where he may have previously thought of the existence of a single force.) In other words *to every action on a given body there is an equal and opposite reaction on some other body.* It follows as a corollary from this law that when two bodies collide their total momentum remains unchanged (see § 27).

Care should be taken to distinguish between (*a*) action and reaction and (*b*) cases of balanced forces. Thus, a book lies at rest on the table. The earth pulls down on the book giving it its weight; the book also pulls up with an equal force on the earth. This is action and reaction. Note that *the action is on the one body, the reaction on the other.* Obviously either force may be called the action, the other being the reaction. The weight of the book is

balanced by the upthrust of the table. Here are two balanced forces. They act on the same body. These two balanced forces have their reactions on a second and third body respectively. It is left to the student to draw a diagram of the book, table, and earth, showing the three pairs of forces and to discuss the equilibrium of each of the three bodies. The forces in case of a ball at the end of a string whirled in a horizontal (also vertical) circle should be analyzed, careful distinction being made between the actions and reactions.

11. The C.G.S. System of Units. — As will be seen in § 13, all mechanical quantities may be expressed in terms of the three fundamental quantities of mass, length, and time. All other quantities which we build up from these fundamental units are called derived units. It is quite arbitrary as to what distance is taken as a unit of length and likewise for a unit of mass and a unit of time. In France, the centimeter, gram, and second, respectively, were selected.

The meter is defined to be the distance between two marks on a platinum bar preserved at the International Bureau of Weights and Measures at Sèvres, near Paris. A one-hundredth part of a meter is called the centimeter. The meter was made as nearly the length of one ten millionth of the quadrant of the earth's meridian passing through Paris as it was possible to determine at the time of the establishment of the unit. Later experiments, when more accurate instruments were available, showed that the quadrant was 10,000,880 meters. No significance is to be attached to the difference. The only advantage that this French unit has over other standards of length is its decimal division.

The gram mass is defined as 1/1000 of the mass of a certain block of platinum kept also at Sèvres and known as the kilogram prototype. The kilogram was intended to be the mass of 1000 cc. of water at 4° C., the temperature of its maximum density. Accurate measurements now show that a kilogram of water at 4° C. occupies 1000.027 cm^3.

Since the liter of volume was originally defined not in terms of the centimeter but as the volume occupied by one kilogram of water at 4° C., it is equal to 1000.027 cm^3. Hence for experiments where an accuracy of 3 parts in 100,000 is demanded, a distinction must be made between the cm^3. and the milli-liter. Since the symbol cc. has been used so interchangeably for the cm^3. and the

thousandth part of a liter, there seems to be some tendency to avoid its use altogether.

The second is defined as 1/86,400 of the mean solar day.

In this system, the unit of force, called the dyne, is defined as that force which will accelerate a mass of one gram at the rate of one centimeter per second, every second. We see that this is a derived unit since it is expressed in terms of the fundamental units of mass, length, and time. Then, because of this manner of choosing the size of the unit force, the constant k in Eq. 20 becomes unity and we have

$$F = ma. \qquad \begin{array}{l} F \text{ in dynes} \\ m \text{ in grams of mass} \\ a \text{ in cm./sec}^2. \end{array} \tag{23}$$

Therefore, in the c.g.s. system, when Eq. 23 is used, F must be expressed in dynes when m is in gms., and a is in cm./sec². If we measure the force of attraction between a gram of matter and the earth, we find that the force varies with the distance from the center of the earth, *i.e.* with altitude. At sea level at the equator, $g = 978.0$ cm./sec². Hence the force of attraction on each gram at that place is 978.0 dynes. At the poles $g = 983.2$. The variation over the earth's surface is about one half of one per cent. Over land bodies at the same altitude, the value of g varies slightly due to geological variations. The very sensitive Eötvös balance is used to locate "salt domes" in undrilled oil territory, careful study being made of the very small variations of the gravitational attraction over the area. Thus the weight of a body varies with its location and so in most accurate work the gravitational force units are avoided. The dyne is independent of the earth's attraction so it is called an absolute unit. When the gravitational force on two grams of matter has to be used, the force F in Eq. 23 is taken as $2\,g$ dynes. For latitudes between 30° and 60°, 980 is a quite accurate value of g.

12. The English (F.P.S.) and Engineering System of Units. — In England the foot, pound, and second were selected as the fundamental units. The foot is defined to be one third of the distance between two marks on a bronze bar, the yard, kept at the Exchequer in London. The pound is defined to be the mass of a certain block of platinum kept also at the Exchequer (the pound mass). But in the engineering system, the gravitational force on

that mass is so much more important than its property of inertia that the word *pound* is restricted to the pound weight.

Let us discuss an experiment to see how much matter must be acted upon by the force of one pound in order that it may be given an acceleration of 1 ft./sec^2. Place a cart on a smooth level table as shown in Fig. 6. We shall make $m_1 = 1$ and $m_2 = 0$. The force on the system is the weight of one pound. Now the cart is loaded until the measured acceleration of the system is just 1 ft./sec^2. Reducing friction to a minimum, the car and contents are found to weigh about 31.2 lbs. Thus a body weighing 32.2 lbs. is given 1 ft./sec^2. acceleration by the force of one pound. A value of g accurate enough for nearly all engineering work at ordinary latitudes is 32.2 ft./sec^2. In order to use Eq. 23 with F in pounds and a in ft./sec^2., m must be replaced by W/g, where W is the weight of the body in pounds. Thus a body weighing 32.2 lbs. is said to have one Engineering Unit of Mass and a man weighing 161 lbs. to have 5 units of mass. This unit of mass is sometimes called the "slug" or the g-pound but no name has been generally accepted. W/g is seen to have the proper units of mass, being a force divided by an acceleration. Furthermore, it has the invariable feature of mass. W and g vary in the same proportions so that W/g is a constant everywhere. We thus have in the Engineering system the formula

$$F = \frac{W}{g} a. \qquad \begin{array}{l} F \text{ in lbs.} \\ \frac{W}{g} \text{ in slugs of mass.} \\ a \text{ in ft./sec}^2. \end{array} \tag{24}$$

It is to be noted that we could use Eq. 24 in the c.g.s. system. If F is to be in gms. (weight), then W will be in gms. and $g = 980$ cm./sec^2. and the unit mass will be one whose weight is 980 gms. Conversely, in the English system if we wished to keep the word *pound* as a unit of mass, we would use Eq. 23. The unit force giving 1 lb. mass an acceleration of 1 ft./sec^2. is called the poundal. Then in Eq. 23, F would have to be expressed in poundals in order to have $k = 1$.

Eq. 23 is generally used for the c.g.s. system of units, while Eq. 24 is nearly always used in engineering work.

In conclusion, we should examine the logical difference in the two systems. In both systems the equation $F = ma$ is to be used.

In the metric system, the units of m and a were arbitrarily selected and therefore the unit of force of a proper size had to be selected so as to make the above equation hold. In the engineering system, however, the selection of the pound weight unit of force (a gravitational pull) and the foot and second pre-determines the units of F and a, and therefore a new unit of mass had to be selected so as to make the equation hold.

PROBLEMS

1. A cord is hung over a pulley. At one end is a mass of 5 kg. At the other end a mass of 7 kg. Find the acceleration of each mass and the tension in the cord.

SOLUTION: The resultant force moving the system is 2000 gms. weight or 1,960,000 dynes. The mass of the system is 12,000 gms. Hence 1,960,000 $= 12{,}000\,a$. $a = 163\frac{1}{3}$ cm./sec². The tension T pulls upward on each mass. Applying Newton's second law to each mass, we have

$$T - 5000 \times 980 = 5000\,a$$
$$7000 \times 980 - T = 7000\,a$$

Adding these equations we get the equation used above. This is an example of the general theorem that for a rigid body all internal forces occur in pairs of equal and opposite forces and hence cannot affect the motion of the body. Only external forces need be considered. Substituting the value of a in either of the above equations, the tension T in dynes may be found.

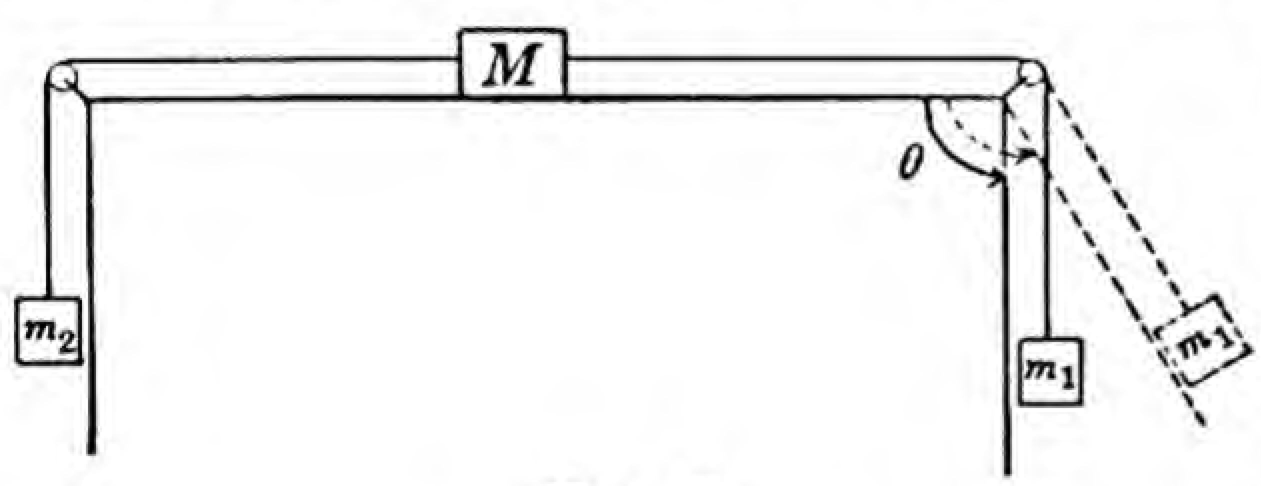


FIG. 6

2. In Fig. 6, $M = 1$ kg., $m_1 = 500$ gms., and $m_2 = 200$ gms. Find acceleration of the "system": (*a*) when $\theta = 90°$, (*b*) when $\theta = 120°$. Three seconds after starting, the string is cut just above m_1. In each case find how far M will go in the next five seconds.

3. Find the tension in each string in Problem 2 (*a*).

4. A mass of 1 kg. at rest is acted upon by a force of 3000 dynes for 3 minutes. No other force is acting. How far will the body go in that time, and what velocity will it acquire?

5. (*a*) A wagon is traveling at a constant velocity on a horizontal road. Describe with vector diagram the velocity of a point on the rim of the wheel for each 45° of rotation of the wheel during a complete revolution. Consider the velocity as made up of two components, one of translation and one of rotation. Likewise describe the acceleration for each 45° of rotation.

(*b*) Solve also for the case when the wagon is uniformly accelerated.

6. An elevator starting from rest reaches full speed of 5 ft. per sec. after moving 10 ft. With what force will a man "weighing" 150 lbs. push down on the floor: (*a*) when the elevator is starting upward, (*b*) when it is starting downward? Derive the general formula for such a problem.

7. Two buckets, each weighing 800 lbs., are suspended from the ends of a rope passing over a windlass. Twenty pounds of water is poured into one of the buckets. Find how far it will descend in 10 seconds, neglecting friction.

8. (*a*) A certain train weighs 100 tons. Starting from rest it reaches a speed of 30 miles per hour after moving 6600 ft. If the acceleration is uniform, what is the average pull of the locomotive? Assume that friction is constant and equals 10 lbs. per ton.

(*b*) The above train, moving at a speed of 30 miles per hour, is brought to a standstill in 16 seconds. Find the additional friction force required.

9. Solve Problem 8 for a train moving up a $\frac{1}{2}$ per cent grade ($\tan \theta = 0.005$).

10. A locomotive is traveling on a curve. If the radius of the curve is r, the velocity v, and the mass m, find the angle, θ, of inclination such that no outward force is exerted against the rails by the flanges of the wheels. Show that the problem may be solved by considering either the actions or the reactions.

11. A railroad curve of 800 ft. radius is to be banked so that a 100-ton locomotive traveling 30 miles/hr. will just not push outward on the rails. If the track width is 4 ft. $8\frac{1}{2}$ in., what elevation must be given the outer rail? What force would the track have to exert upon the train if the road bed were not banked?

12. In the case of the system shown in Fig. 7, the planes are frictionless. Find the direction and magnitude of the acceleration. What frictional force would cause the velocity to be constant?

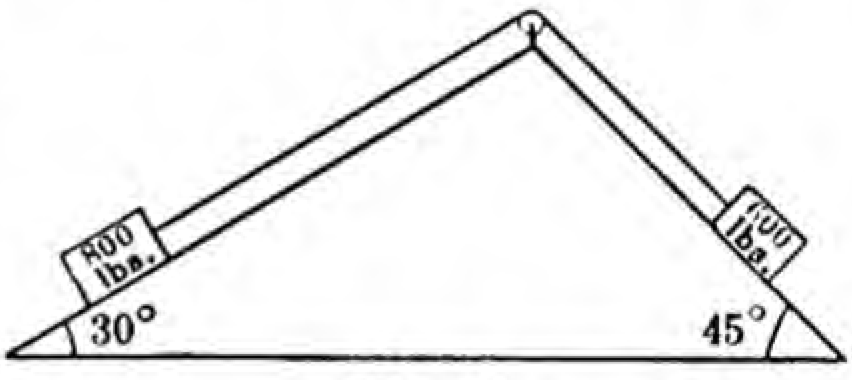


FIG. 7

13. In Problem 2, part (*a*), let m_1 be a bucket weighing 200 gms. and filled with 1000 gms. of water. At the instant of starting of the motion of the system, let the bucket start to leak at a constant rate so that it is empty at the end of five seconds. Evaluate

$$\int_0^t F\,dt.$$

14. A string will stand a tension of no more than 50 lbs. without breaking. What is the maximum number of times per minute a 2-lb. weight can go around a circle of 2 ft. radius if it moves on a smooth horizontal surface and is constrained to the circle by the string?

13. Dimensions. — As mentioned before, all of the physical quantities used in mechanics may be expressed in terms of the three fundamental units of length, mass, and time. *The dimensions of a physical quantity are the powers of the fundamental units from which it is derived.* For instance, a velocity unit must express a length divided by a time. Its dimensions are (+ 1) in length, and (− 1) in time. Acceleration is expressed in centimeters per second per second and its dimensions are (+ 1) in length, (− 2) in time. The

dimensional formula for velocity is $[L]\ [T]^{-1}$, for acceleration $[L]\,[T]^{-2}$. From Eq. 20 we see that the dimensional formula for force is $[M]\,[L]\,[T]^{-2}$, k being merely a number and having no dimensions.

Dimensions are convenient in specifying a derived unit which has no name; thus, both velocity and acceleration are expressed in terms of their dimensions, vel. as cms./sec., acceleration as cms./sec^2., etc. Dimensions afford a valuable method of checking the accuracy of an equation. All terms in an equation must be of the same dimensions. Thus, suppose we have a body of mass m moving vertically upward with a velocity u. A force F is applied upward on the body until it has moved through a distance h. It then has a velocity v. We write for the final kinetic energy of the body,

$$\tfrac{1}{2}\, mv^2 = \tfrac{1}{2}\, mu^2 + Fh - mgh.$$

The student should show that each term of this equation has the dimensions ML^2T^{-2}. The equation is, therefore, dimensionally correct. Every correct equation must satisfy such a relation.

We are sometimes enabled to determine the relation between physical quantities by examining the dimensions of the quantities. Thus, suppose we know from experiment that the velocity of sound in an elastic medium depends upon the density δ and the elasticity of the medium and upon no other factors. Density is mass per unit volume; the modulus of elasticity involved in a compressional sound wave is the bulk modulus, β, which is force per unit area divided by change in volume per unit volume (Eq. 39). The equation may be generally stated as

$$v = k\delta^x\beta^y,$$

where k is a proportionality constant, x and y are the unknown exponents of the density and the modulus of elasticity. Expanding into fundamental units,

$$\begin{aligned} LT^{-1} &= k(ML^{-3})^x(ML^{-1}T^{-2})^y \\ &= kM^{x+y}L^{-3x-y}T^{-2y}. \end{aligned}$$

If this equation is to hold true, then the dimensions of M must be the same on each side and likewise for L and T.

$$\begin{aligned} &\text{For } M, & 0 &= x + y. \\ &\text{For } L, & 1 &= -\,3\,x - y. \\ &\text{For } T, & -\,1 &= -\,2\,y. \end{aligned}$$

From these equations we obtain

$$\begin{aligned} x &= -\,\tfrac{1}{2} \\ \text{and } y &= +\,\tfrac{1}{2}. \end{aligned}$$

Hence the preceding equation becomes

$$v = k\sqrt{\frac{\beta}{\delta}}.$$

It is seen that dimensional reasoning will not give the value of k. Had not all the proper quantities been considered which affect the velocity, then the three equations above would not have been simultaneous and would not have been satisfied by single values of x and y. Thus, had we assumed, as might be thought possible for a gas, that the velocity depended upon only the atmospheric pressure and the elasticity (bulk modulus, β), the following equations would have been obtained:

$$\begin{aligned} 0 &= x + y, \\ 1 &= -x - y, \\ -1 &= -2x - 2y, \end{aligned}$$

which are entirely inconsistent.

PROBLEM

Test the following statements to see if they are dimensionally consistent: (*a*) the difference in pressure between the gas inside and outside of a soap bubble depends only on the surface tension (force per unit length) of the soap film and the radius of the bubble; (*b*) the period of vibration of a simple pendulum depends only on the length of the pendulum and on the acceleration of gravity; (*c*) the velocity of a wave along the cord depends only on the tension in the cord and the mass per unit length of the cord.

EQUILIBRIUM

14. The General Conditions for Equilibrium of a Rigid Body. — A particle is a body whose dimensions are infinitesimal. A body is said to be rigid if it has an appreciable size and if all the distances between its particles remain unchanged no matter what forces are applied to it. Although there are no perfectly rigid bodies, many solids are sufficiently undeformed by even large forces that they are practically rigid.

A rigid body is said to be in equilibrium when it has no acceleration of either translation or rotation.

To assure no acceleration of translation we must have — ***Condition I*** — *The vector sum of all external forces acting on the body must be zero.* To assure no acceleration of rotation we must have — ***Condition II*** — *The vector sum of the moments of all the external forces must be zero about any point whatsoever.*

These are general statements for the equilibrium of a rigid body which is acted upon by a set of forces distributed in space. Condition I states that the force-vector polygon, even though a skewed figure not all in one plane, is a closed figure.

Let us consider next the moments of the forces with reference to axes passing through some point, O, in space. Each force has a tendency to produce rotation about an axis through O which is perpendicular to the plane containing the point O and the force vector. The moment of the force, $F \times l$ (§ 2) is represented by a vector of length $Fs \sin \theta$ perpendicular to that plane. Likewise, the moments of the other forces about their appropriate axes through O are definite vectors. These various moment-of-force vectors when placed so as to form a vector polygon in space must also form a closed figure if the body is in rotational equilibrium. This idea is concisely stated in Condition II. Although the moment of a force always involves the tendency of a force to produce rotation about an *axis*, there should be no ambiguity if we use the shorter expression, "the moment of the force about a *point*." Without this somewhat common usage of the word *point*, Condition II when applied to a three-dimensional configuration of forces would be a much more involved expression.

Discussion in this text will be limited to co-planar forces or forces acting all in the same plane. If we call the forces F_1, F_2, $\ldots F_n$, their x and y components $X_1, X_2, \ldots X_n, Y_1, Y_2, \ldots Y_n$, respectively and their lever arms about any point O whatsoever as $d_1, d_2, \ldots d_n$, respectively, then the above conditions reduce to the following algebraic equations.

$$X_1 + X_2 + \cdots X_n = 0, \quad Y_1 + Y_2 + \cdots Y_n = 0, \tag{25}$$

and

$$F_1d_1 + F_2d_2 + \cdots F_nd_n = 0. \tag{26}$$

To show that Eq. 25 is a statement of the two-dimensional case of Condition I, consider Fig. 8. The forces 1 to 5 satisfy Condition I since the vector polygon is closed. Project all of these forces onto the two axes in order to find the values of the X's and Y's. It is seen that the vector AB equals $X_1 + X_2 + X_3$. The vector $(-BA)$ equals $X_4 + X_5$.

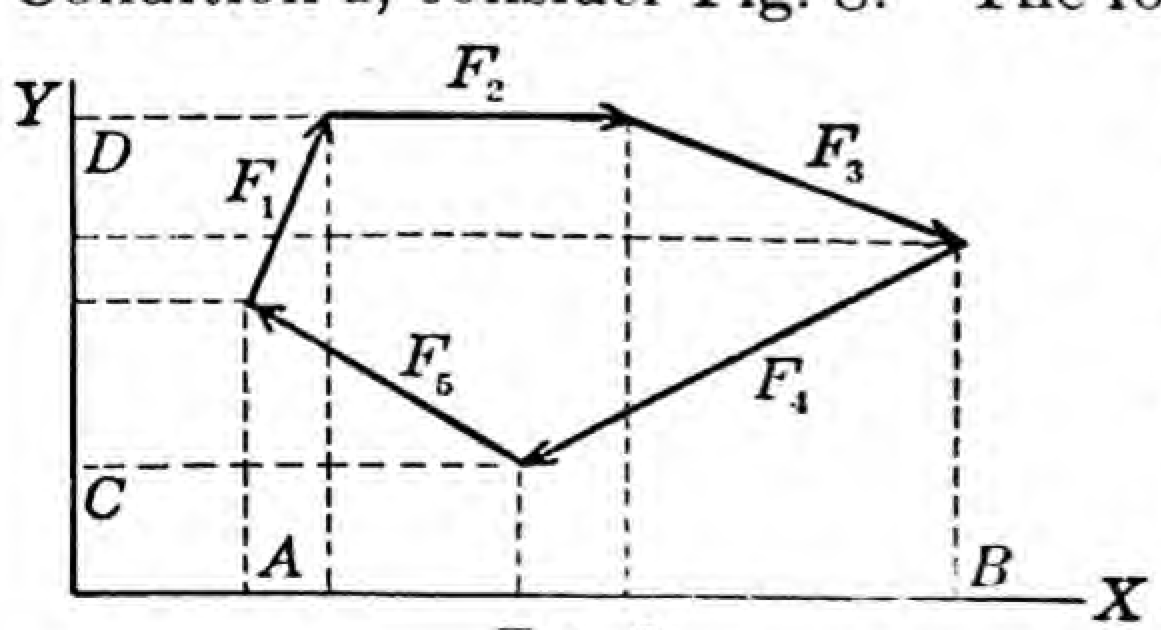


FIG. 8

Hence, algebraically,

$$X_1 + X_2 + X_3 + X_4 + X_5 = 0.$$

Likewise for the Y values related to the vector CD. Eq. 26 is a mere algebraic statement of Condition II since only one axis is involved and the moments have + or − directions along that line.

A particle is in equilibrium when it has no linear acceleration. *Condition I (or its equivalent in two dimensions of Eq. 25) is a sufficient condition for equilibrium of a particle.* Condition II does not apply since there is no meaning to the rotation of a particle.

When a set of external forces is applied to a rigid body, stresses are set up inside the body opposing the forces which slightly distort the body. It may be shown that the set of internal stresses are at all times in equilibrium among themselves and so they do not have to be considered in so far as the equilibrium of the whole body is concerned. If a rigid body is in equilibrium, then all of its particles are in equilibrium also, under a combination of external and internal forces. If any single one of these forces at a point is unknown, it may be found by applying only Condition I.

Sometimes we know all the forces and their points of application. We may then evaluate the left-hand members of Eqs. 25 and 26 and see if the body can be in equilibrium. At other times we may know that a body is in equilibrium and apply the equations in order to solve for some unknown force or lever arm.

EXAMPLES

The board in Fig. 9 has one-inch rulings. We wish to see if the board is in equilibrium and if it is not, to find what single force will keep it in

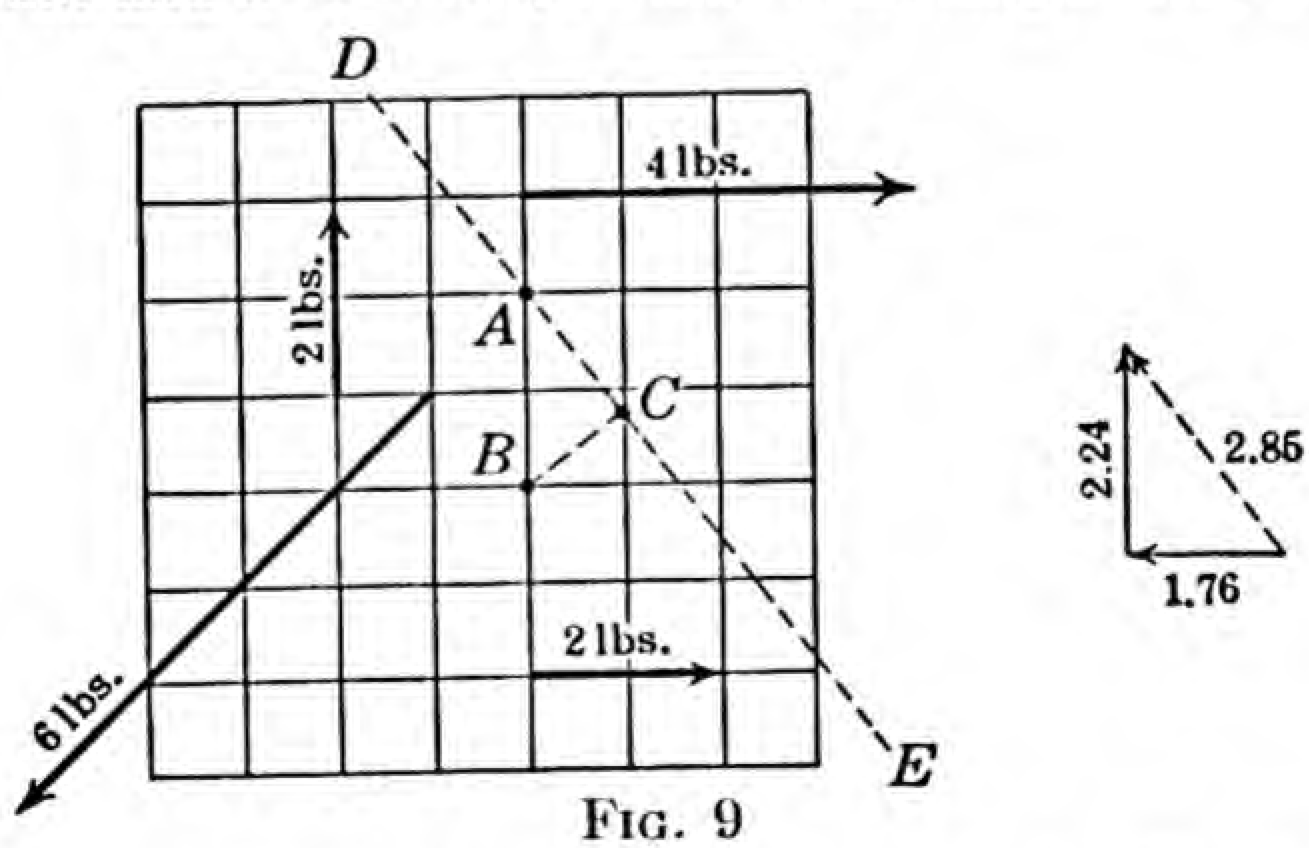


FIG. 9

equilibrium and where that force must be applied. Eq. 25 will first be applied.

$$2 + 4 - 6 \cos 45° = + 1.757 \neq 0$$
$$2 - 6 \sin 45° = - 2.243 \neq 0.$$

Therefore the board will have an acceleration of translation. To produce equilibrium additional x forces of $- 1.757$ lbs. and y forces of $+ 2.243$ lbs. must be applied. A single force which would produce this effect is

$$\sqrt{1.757^2 + 2.243^2} = 2.849 \text{ lbs.}$$

The direction of the force is shown by the dotted line of the force-vector triangle in Fig. 9.

Applying Eq. 26 for the point B,

$$(4 \times 3) + (2 \times 2) - (6 \times 1.414) - (2 \times 2) = + 3.515 \neq 0.$$

To produce rotational equilibrium a counter-clockwise moment of $- 3.515$ lb. in. about point B must be applied. In order that this moment be produced by the required force of 2.849 lbs., the lever arm must be 3.515/2.849 = 1.234 in. Therefore the single force of 2.849 lbs. must be applied along the line ED parallel to the hypotenuse of the force-vector triangle and must be directed toward D. The perpendicular distance from B to the line DE must be BC = 1.234 in.

If, by chance, Eq. 26 had been applied for the point A, we would have obtained

$$(4 \times 1) + (2 \times 2) + (6 \times 0) - (2 \times 4) = 0.$$

Since the force of 2.849 lbs. must still be applied, we must conclude that it cannot produce any moment of force about A or else the sum of the moments could not still be zero. Therefore we conclude that the equilibrating force must pass through A. It is thus seen that Condition II by itself is not sufficient to insure equilibrium. The student should show likewise how Condition I by itself is not sufficient.

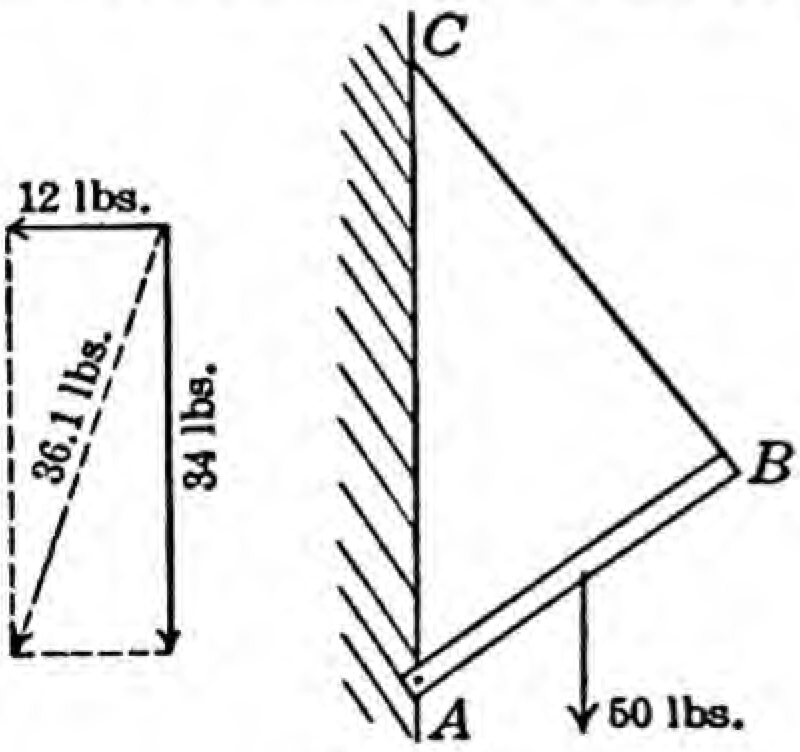


Fig. 10

Now consider the case of a 6 ft. rod AB (Fig. 10) weighing 50 lbs., pivoted at A, and supported at B by a flexible rope from C, of 8 ft. length. The rope makes a right angle with the rod. As we shall see in the treatment of Center of Mass, the weight of the bar may be considered as acting at its center of gravity. We may first find the tension T in the rope. In such a system known to be in equilibrium we find it convenient to take as a center of moments a point where an unknown force acts; then the moment arm of that force is zero so that it does not appear in the equation. So taking moments about A, we eliminate the unknown reaction of the wall on the rod. The lever arm of the weight of the bar with reference to A is found to be 2.4 ft. $T \times 6 - 50 \times 2.4 = 0$. So $T = 20$ lbs. Now knowing two of the three forces acting on the bar, we may find the components of the reaction of the wall by applying Condition II. The vertical component of T may be found to be 16 lbs. and the horizontal

component to be 12 lbs. Calling the components of the wall thrust X_1 and Y_1 and the force itself F_1,

$$-50 + 16 + Y_1 = 0, \qquad Y_1 = 34 \text{ lbs. upward.}$$
$$-12 + X_1 = 0, \qquad X_1 = 12 \text{ lbs. outward from wall.}$$
$$F_1 = \sqrt{12^2 + 34^2} = 36.1 \text{ lbs.}$$

The push of the rod on the wall must be equal and opposite to F_1 and is shown on the figure.

PROBLEMS

1. A one-ton girder, 30 ft. long, leans against the wall with the bottom end 5 ft. from the wall. Find the components of the thrust and the total thrust at each end of the girder. Consider the wall smooth, *i.e.* let no friction forces exist. Then the only reaction at the wall is an outward and inward force. What is the friction force between the girder and the ground?

2. A man weighing 150 lbs. stands at the middle of a slack wire and it sags 2 ft. below the tops of the 4 ft. vertical posts which support it. The posts are 16 ft. apart. Guy wires fastened from the tops of the posts fasten to the floor 3 ft. from the foot of the posts. Find the tension in the wire, the tensions in the guy wires, and the compression of the posts.

3. An automobile wheel is 30 inches in diameter. What thrust on the ground must the rear wheels make in order to have the front wheels start over a curb 6 in. high? Consider the weight on the front wheels to be 1500 lbs. If the force were applied by a rope around the axle, what is the least force which would be required and at what angle to the horizontal must it be applied?

4. Determine the forces of action and reaction at each joint of the following structures, neglecting the weights of the members.

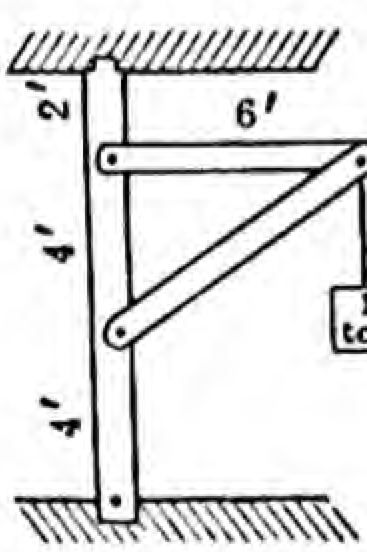


FIG. 11

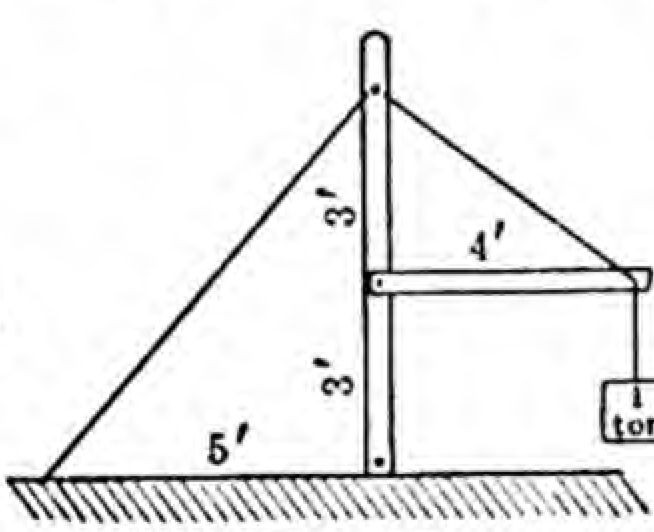


FIG. 12

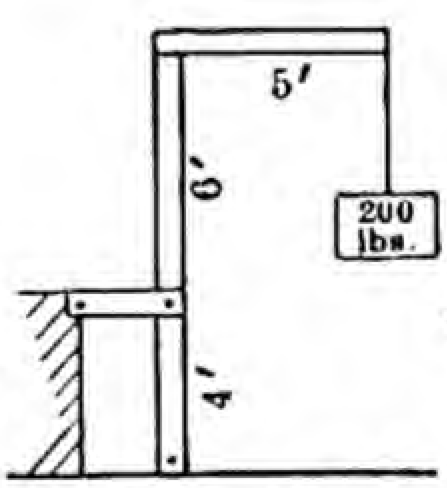


FIG. 13

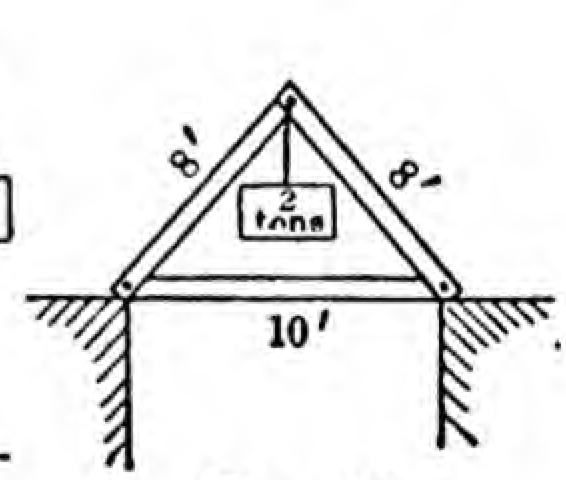


FIG. 14

WORK AND ENERGY

15. Work. — If a force is applied to a body but no motion results (due to equilibrium produced by an equal and opposite force), the force is said to do no work. If the body is in motion but no component of the displacement is in the direction of the force, still the force is said to do no work. Only when the body has a displacement in the direction of the force is work said to be done by the force. *It is agreed to measure the work by the product*

of the force and the component of the displacement which is parallel to the force. As explained in § 2, we may then write, for an infinitesimal displacement,

$$dw = F\,ds\cos\theta. \tag{27}$$

From this equation it is seen that we may also measure work by the product of the displacement and the component of the force which is parallel to the displacement.

Eq. 27 may be written,

$$F\cos\theta = \frac{dw}{ds},$$

which states that the derivative of work with respect to the displacement gives the component of the force which is parallel to the displacement. So in order to find w we must find the quantity which, when differentiated with respect to s, gives $F\cos\theta$. That quantity is the integral of $F\cos\theta$ with respect to s.

$$w = \int F\cos\theta\,ds. \tag{28}$$

This expression becomes integrable either by expressing $F\cos\theta$ as a function of s, as is done in § 18, or by changing both $F\cos\theta$ and ds to a common new variable, as in § 19.

The c.g.s. unit of work is *the work done by a force of one dyne acting through a distance of one centimeter.* This unit is called the *erg.* The *joule* is 10^7 ergs.

In the English system the unit of work is the *foot-pound, the work done by a force of one pound acting through the distance of one foot.*

16. Power. — Power is the time rate of doing work. In mathematical notation,

$$P = \frac{dw}{dt}. \tag{29}$$

By combination of Eqs. 27 and 29, we find

$$P = F\cos\theta\frac{ds}{dt} = F\cos\theta\cdot v. \tag{30}$$

If both the velocity and the component of the force are variable, this expression gives only the instantaneous power at the time when F, θ, and v have the values given.

The *c.g.s. unit of power* is the erg per second. A more convenient unit is the *watt, which is the joule per second.* Still another unit is the *kilowatt* (1000 watts). A common unit of work in measuring

electrical energy is the *kilowatt-hour*. It is the work done (or energy consumed) in one hour when the rate is a kilowatt (1000 joules every second) and is therefore equal to 3.6 million joules. The English unit of power is the *horsepower*, which is 550 foot-pounds per second. The student is expected to determine the units in which the power is expressed (in Eq. 30) when various units for F and v are used.

17. The Graphical Representation of Work. — If the component of the force which is parallel to the displacement is plotted against the displacement s, it may be shown that the area under the curve on such a diagram gives work. Suppose the force has the constant value F_1 (Fig. 15), and the body travels a distance s_1 in the direction of the force, then the work done, being $F_1 \times s_1$, is obviously the area under the force-distance line. If the force is variable (Fig. 16) and we wish to find the work when the body moves from s_1 to s_2, it will now be necessary to divide the distance into infinitesimal portions of length ds through which the force varies by only an infinitesimal amount. In such an infinitesimal range, the work is dw and $dw = F\,ds$, which is the area of the unshaded strip. The summation of all the strips obviously gives the total work between the limits, s_1 and s_2, and this is the area under the force-distance curve. We have seen (Eq. 28) that the total work in the case of a variable force is

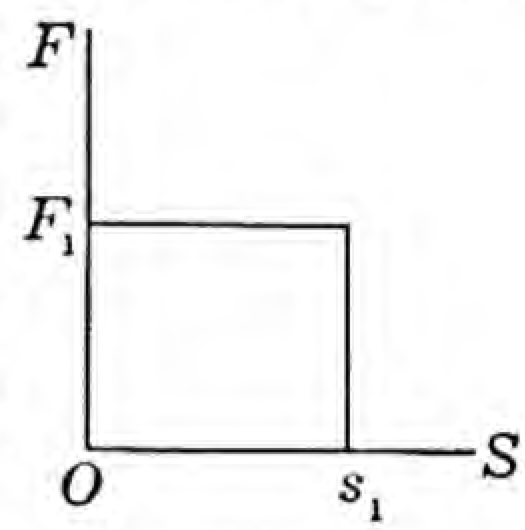


FIG. 15

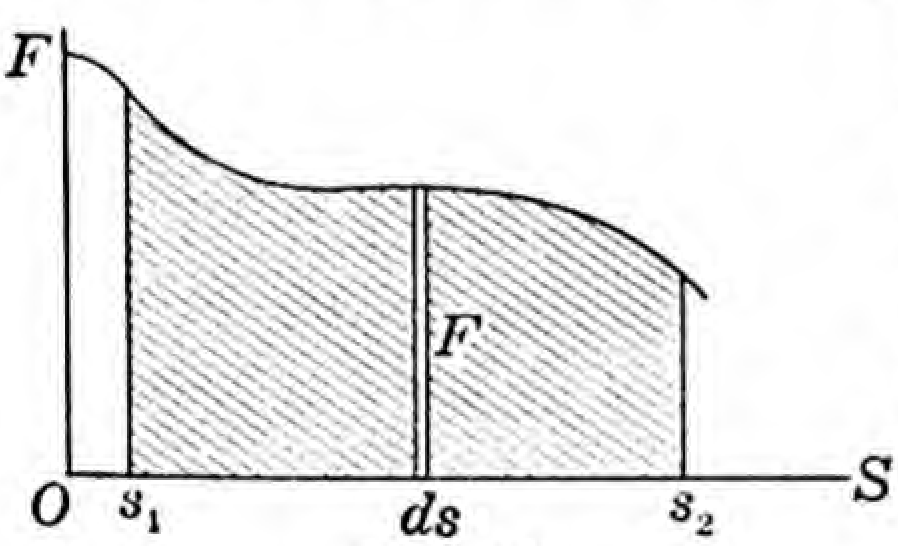


FIG. 16

$$w = \int F\,ds.$$

Hence the process of integration in this case is a process of summation.

18. The Stretched Spring. An Illustration. — Let us consider as an example of the above process the work done in stretching a spring. The applied force F must at each instant equal the elastic return force F' and must be in the opposite direction to it. By Hooke's law, the applied force must be proportional to the amount of stretch s. Therefore, $F = cs$. The value of c is called the

spring constant and is the force required to stretch the spring one unit of length. Fig. 17 shows graphically the relation $F = cs$.

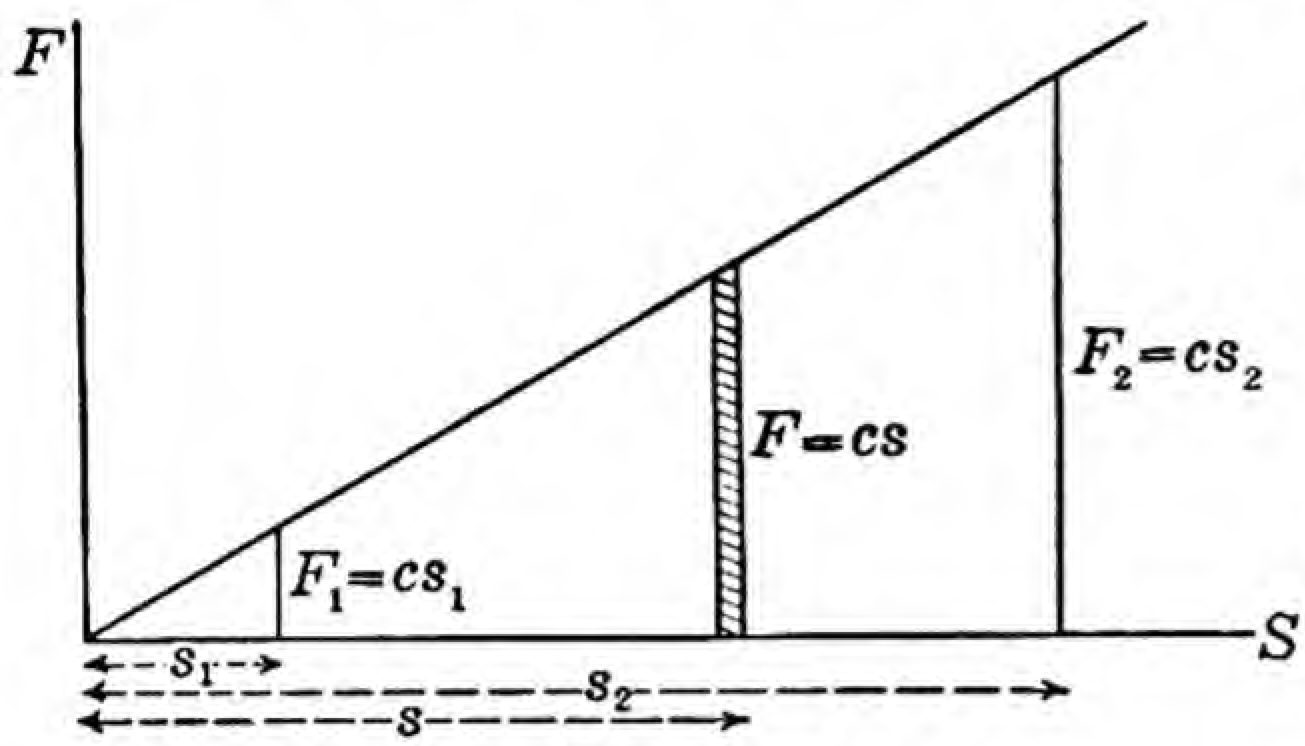


FIG. 17

Substituting the value of F in Eq. 28 and noting that $\theta = 0$, we have

$$w = \int cs\, ds. \tag{31}$$

The integral of this (the quantity which when differentiated with respect to s gives cs) is

$$w = \tfrac{1}{2}\, cs^2 + c', \tag{32}$$

where c' is the constant of integration. If we choose to start measuring work when the displacement is zero, that is, take $w = 0$ when $s = 0$, then $c' = 0$. For this case Eq. 32 then becomes

$$w = \tfrac{1}{2}\, cs^2. \tag{33}$$

Inspection of Fig. 17 shows that the area under the curve out to the indefinite value s is equal to $\frac{1}{2}\, cs^2$. The area out to the value of s_1 is $\frac{1}{2}\, cs_1^2$ and out to s_2 is $\frac{1}{2}\, cs_2^2$. The area under the curve between the values of s_1 and s_2 is

$$\tfrac{1}{2}\, cs_2^2 - \tfrac{1}{2}\, cs_1^2, \tag{34}$$

and this area represents numerically the amount of work done in stretching the spring through the displacement $s_2 - s_1$. In this special case we see that the area under the curve between two ordinates is obtained by taking the indefinite integral and then finding the difference in the values of this integral expression for the two definite ordinates considered. In fact, this is the general case, as is proved in Integral Calculus. It is customary to indicate these processes by the following notation.

$$W = \int_{s_1}^{s_2} cs\, ds = \tfrac{1}{2}\, cs^2 \Big]_{s_1}^{s_2} = \tfrac{1}{2}\, cs_2^2 - \tfrac{1}{2}\, cs_1^2. \tag{35}$$

When F and ds have the same direction, the work is done *by the force on the body* on which the force acts and the product $F\, ds$ and the integral $\int F\, ds$ will be positive. When F and ds have opposite directions, the body moves oppositely to the direction of the force, the work is done *by the body against the force*, and the expression for the work gives a negative value.

19. Potential and Kinetic Energy. — Energy may be defined as the capacity to do work. When a force acts upon a body and does work upon it, (1) the momentum of the body may be changed; (2) the body may be moved in opposition to return or restoring forces; (3) the body may be moved in opposition to frictional forces which do not tend to return the body to its initial position. In the first case the body may give up a portion or all of its momentum when it encounters opposition to its motion and may, therefore, exert a force (equal to the rate of change of momentum) and do work upon an opposing body. *The energy a body possesses by virtue of its motion is called kinetic energy.* In the second case when work is done against restoring forces, these restoring forces, when the applied force is removed, will return the body to its initial position and thus do work equal to that done upon the body. In this case the ability to do work is potential. This energy is stored up in the system. *A body possessing energy by virtue of its position or the configuration of its parts is said to have potential energy.* The stretched spring has potential energy. A body lifted against gravity to a certain height has potential energy of position. It exerts a force due to the pull of gravity upon it and it possesses the ability to do work upon returning to a lower level. In the third case, although work is done upon the body, the body does not acquire the capacity to do work. The energy is not stored in the body either as kinetic or potential energy, but has been dissipated as heat.

Experiment shows that the energy acquired by a body (if none has been dissipated as heat) is equal to the work done upon it. When energy is acquired by one body, experiment shows that an equal amount must have been lost by the body through which the force was exerted. *Energy may be converted from one form to another, but experiment shows that it cannot be created or destroyed. This statement is known as the Law of the Conservation of Energy.* For example, imagine a moving body suddenly attached to the end of a spring; the spring will be stretched and the body brought to rest. The kinetic energy of the moving body has been converted into potential energy of the stretched spring. Consider a projectile fired vertically upward with a certain initial velocity. After a certain time the upward velocity will be reduced to zero. At the instant when the velocity is zero the body will have potential energy equal to the kinetic energy lost.

We shall now obtain an expression for the amount of kinetic energy possessed by a moving body.

Let a mass m, which is at rest, be acted upon by a force F (which may be variable). From §§ 2 and 15, we have the integral expression for the work done by the force, namely,

$$W = \int_0^s F \cos \theta \, ds.$$

The component $F \cos \theta$ is parallel to the displacement and produces an acceleration a parallel to it. Therefore a, dv, and ds are all in the same straight line and may be treated algebraically. We have $a = dv/dt$ and $v = ds/dt$. Hence

$$\begin{aligned} W &= \int_0^s F \cos \theta \, ds = \int_0^s ma \, ds = \int_0^s m \frac{dv}{dt} ds \\ &= m \int_0^v v \, dv = \tfrac{1}{2} mv^2. \end{aligned} \tag{36}$$

The above expression gives the work done on the body by any force whatsoever. In coming to rest the body may do an amount of work equal to $\frac{1}{2} mv^2$, hence this expression represents the kinetic energy of the body.

If the body were already in motion with a velocity v_1 when the force was applied and if it attained a final velocity v_2, the integration in Eq. 36 would have been performed between the limits v_1 and v_2. Then the work done would have been

$$W = \int_{v_1}^{v_2} mv \, dv = \tfrac{1}{2} mv_2^2 - \tfrac{1}{2} mv_1^2, \tag{37}$$

which represents the increase in kinetic energy produced by the force.

The student is expected to follow through each step of the proof and show the units in which m and v must be expressed in order to obtain energy in ergs and in ft. lbs.

PROBLEMS

1. If one pushes with a constant force on a body which is moving initially at the rate of v_0 feet per second, will more or less work be done during the first second than when the body is moving initially at the rate of $2 v_0$ feet per second? If so, how much? How much work is done in t seconds in each case?

2. A body weighing 20 lbs., starting from rest, slides 100 feet down a 10 per cent grade in 10 seconds. How much energy has been converted into heat through friction?

3. Water flows into a mine 800 ft. deep at the rate of 50 cu. ft. per minute. Find the H. P. of a pump to be used to keep the mine dry. One cubic foot of water weighs 62.4 lbs.

4. A ton of coal is hauled in a car of 400 lbs. out of a vertical shaft 200 ft. deep. During the first half of the journey it is accelerated 4 ft./sec^2. and during the last half, retarded by the same amount. Neglecting friction, find the force exerted, the work done, the average power, and the maximum power. Would the answers be different if the shaft were inclined, the total depth being the same? How would friction affect these answers?

5. A test tube of negligible weight has a cross section of 5 sq. cm. How much work is done in pushing it vertically downward into water to the depth of 15 cm.? Into mercury? In each case how much net work is done on or by the test tube if it weighs 50 gms.?

6. Prove: 1 H.P. = 746 watts and 1 joule = 0.738 ft. lbs.

7. A train starts from rest on a level track and moves through 1200 feet in the first minute. It then begins to ascend a uniform incline, up which it is found to run with constant velocity. Find the inclination of this portion of the track on the supposition that the engine exerts a constant pull.

8. What is the least angular velocity necessary to swing a pail of water weighing 20 lbs. in a vertical circle 4 ft. in diameter, without spilling the water? As the pail descends from the highest point it gains speed due to its loss of potential energy and gain in kinetic energy. What will be the force on the handle when the pail is 30 degrees from the top? At the bottom?

Elasticity

20. Stress and Strain. — An *elastic body* is one in which a distortion is opposed by molecular forces which tend to restore the body to its original shape or volume.

There are three main ways in which we may change the shape of a body. We may apply a force so as to pull apart or push closer together parallel layers of atoms of the body; a pressure may be applied on all sides of a body as hydrostatic pressure and thus decrease in all directions the distances between the atoms; or we may apply a force so that each successive layer of molecules is slid a short distance over the preceding one. In all of these cases we can pick out the internal restoring forces and the areas over which they act. *The internal restoring force per unit area is called the stress.* In each of these three kinds of distortion the change in shape occurs in a different manner. The actual amount of distortion depends on the size and shape of the body and is measured differently in each case, as will now be explained.

The amount of stretch of a wire, for a given applied force, varies directly with the length of the wire. However, the amount of

stretch of each unit length is a constant. *In case of a compression or a stretch, the change in length of each unit length is called the longitudinal strain.*

Similarly, when a given hydrostatic pressure is applied to a body, the change in volume varies directly as the volume of the body, although the change in volume of each unit volume is the same throughout the body. *In the case of change in bulk, the change in volume per unit volume of the body is called the bulk strain.*

In the case of a shear (see Fig. 18) for a given applied force, the amount of shear varies directly as the distance between the two sliding areas considered. *The amount of slide, or shear, between two layers a unit distance apart is called the shear strain.*

In general, we may say that the *strain is defined as the relative amount of distortion or as the fractional change in shape of a body.*

Hooke's Law. — Hooke discovered experimentally that *within what is called the elastic limit, the stress is proportional to the strain.* The ratio of the stress to the strain is called the *modulus of elasticity.*

21. Moduli of Elasticity. — Whenever an external force F is applied to stretch or compress a bar of uniform cross section a, the internal reaction forces act uniformly across the whole cross section. They are parallel to the force F and equal to it. Hence the stress is F/a. The length l will be increased or decreased by an amount e and the strain is defined to be e/l. The ratio of the stress to the strain is known by experiment to be the same for a stretch or a compression, for homogeneous bodies, and is called *Young's Modulus*, Y. Hence,

$$Y = \frac{F}{a} \div \frac{e}{l} = \frac{Fl}{ae}. \tag{38}$$

If hydrostatic pressure is applied to a body and causes a volume strain, the internal reaction forces are perpendicular to the surface to which the pressure is applied and these internal reaction forces per unit area are equal to the applied pressure. If the volume V changes by an amount v, then the *Bulk Modulus* is

$$\beta = p \div \frac{v}{V} = \frac{pV}{v}. \tag{39}$$

In both of the above types of deformation, the return forces are perpendicular to the areas over which they act. When a shearing strain is produced, the case is quite different. The layers

of molecules slide over each other so that the volume of the body is not changed appreciably; a couple is required to produce the strain and *the internal return forces F* (Fig. 18) *are parallel to the surfaces on which they act.* The strain is the distance which one layer of molecules moves with respect to some other layer of molecules divided by the distance between the layers. This ratio is practically the same as the angle ϕ (expressed in radians) in Fig. 18, because in solid bodies the angle is extremely small. The *shear modulus, n,* also called *slide modulus* or *rigidity modulus,* is the ratio of stress to strain in case of shear.

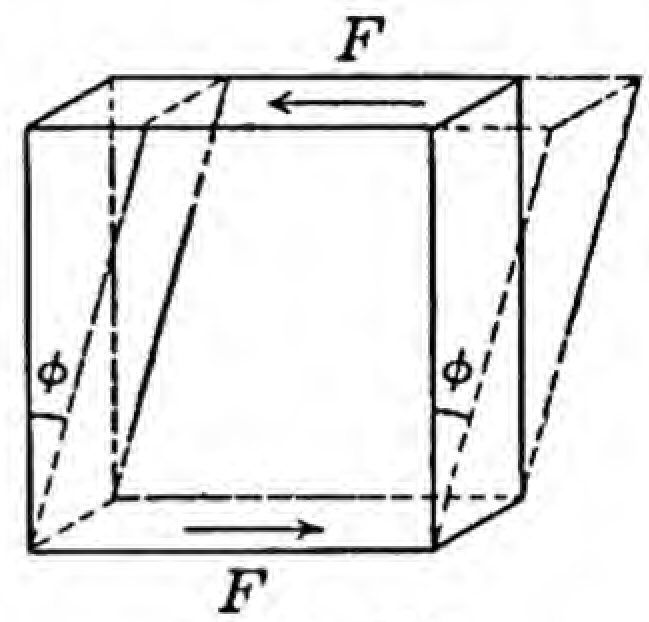


Fig. 18

$$n = \frac{F}{a\phi}. \tag{40}$$

Material	Young's Modulus		Shear Modulus	
	Dynes per cm².	Lbs. per in².	Dynes per cm².	Lbs. per in².
Aluminum	7.0×10^{11}	10.2×10^{6}	2.5×10^{11}	3.63×10^{6}
Brass	9.2	13.4	3.7	5.38
Copper	10.0	14.5	4.2	6.10
Iron, cast	11.5	16.8	5.10	7.41
Iron, drawn	20.0	29.1	8.00	11.6
Steel, mild	22.0	32.0	8.00	11.6
Steel, cast	19.5	28.3	7.5	10.9
Steel, drawn	18.8	27.3		

22. The Twisted Shaft or Wire. — When a wire or shaft is twisted, we notice that areas near the center of any given cross section are not moved as far, with respect to those in another given cross section, as are areas near the edge of the wire. So the return forces per unit area are not constant over the whole cross section. However, the force per unit area is nearly constant over any portion of a very narrow ring whose center is the center of the cross section of the wire. In order to connect the stress with the measurable externally applied couple, we must calculate the return couple due to each cylinder whose cross section is a ring of infinitesimal thickness and add these couples so as to include the whole cross section of the wire. We may then equate this sum to the applied external couple since we observe that the wire is in equi-

librium in its twisted shape. In Fig. 19 let l be the length of the bar, R its radius, and L the moment of the applied couple. In

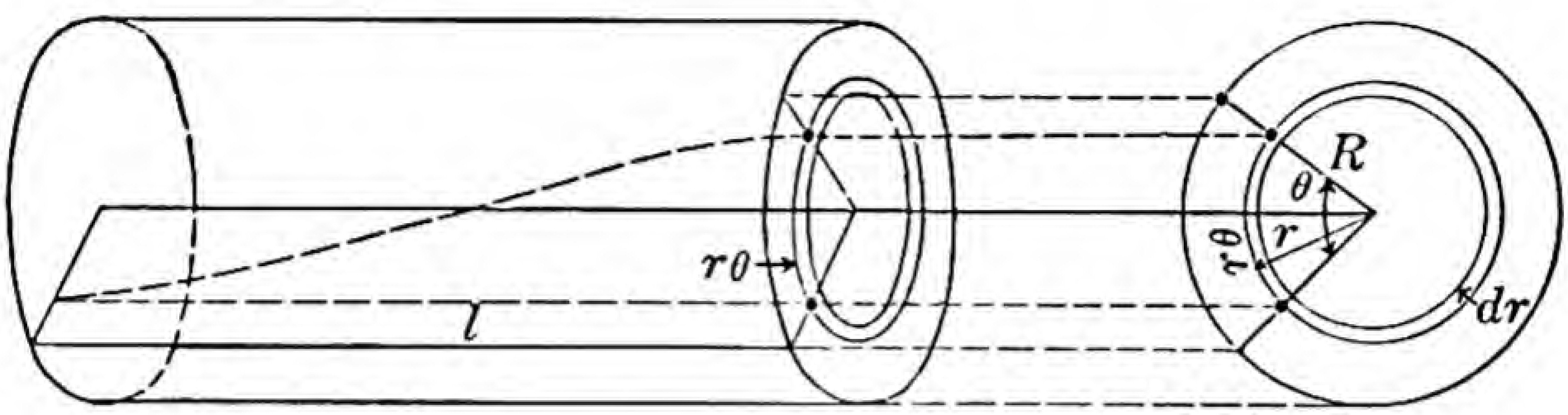


FIG. 19

the cross section at the right-hand end, consider a ring of radius r and thickness dr. The strain is $r\theta/l$ and from Eq. 40,

$$\text{stress} = \frac{dF}{da} = n \times \text{strain} = n\frac{r\theta}{l}.$$

The area of the ring da is $2\,\pi r\,dr$, so the infinitesimal return force on the surface of the ring is

$$dF = \frac{2\,\pi n\theta r^2\,dr}{l}.$$

The lever arm of this force is r. Therefore the infinitesimal moment of the return force, which is equal and opposite to the infinitesimal couple required to twist the cylinder, is

$$dL = \frac{2\,\pi n\theta r^3\,dr}{l}.$$

The sum of all these couples is got by integrating this expression for all values of r from $r = 0$ to $r = R$. The twisted bar is in equilibrium in its strained condition so the sum of the internal couples must equal the applied external couple.

$$L = \frac{2\,\pi n\theta}{l}\int_0^R r^3\,dr = \frac{\pi n\theta R^4}{2\,l}.$$

whence
$$n = \frac{2\,lL}{\pi\theta R^4} \tag{41}$$

in which l, L, θ, and R are easily measurable quantities. The usual manner of determining the modulus of rigidity is to measure these four quantities on a twisted cylinder.

23. Bending Beams. — If a beam of rectangular cross section is supported at the ends and loaded in the middle, it is distorted so

that the upper parts of the beam are compressed and those in the lower part are stretched (Fig. 20). So it is expected that the relation between the sag of the middle part of the beam e would involve the load F, the dimensions of the bar, and Young's Modulus. For small bends, the bar takes on the shape of a segment of an annulus. The formula for the sag e, of the center, which will not be derived, is

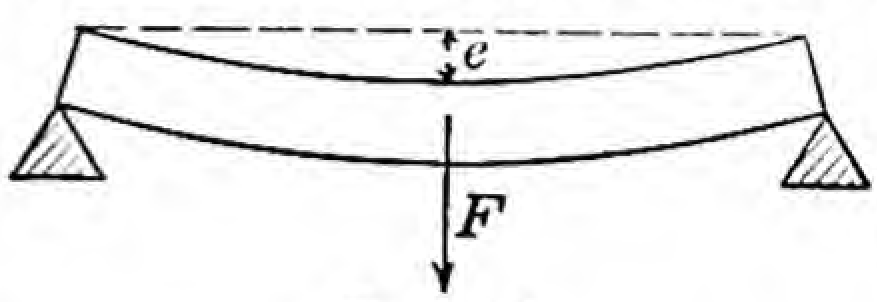


FIG. 20

$$e = \frac{Fl^3}{4\,Ybd^3}, \qquad (42)$$

where l is the length of the bar, b the breadth, and d the depth.

24. Limitations of Hooke's Law. — Hooke's law is strictly true for only very small strains. Typical tension-strain relations for several materials are shown in Figs. 21 and 22. Fig. 21 gives an enlarged view of the early portion of the curves in Fig. 22. In the

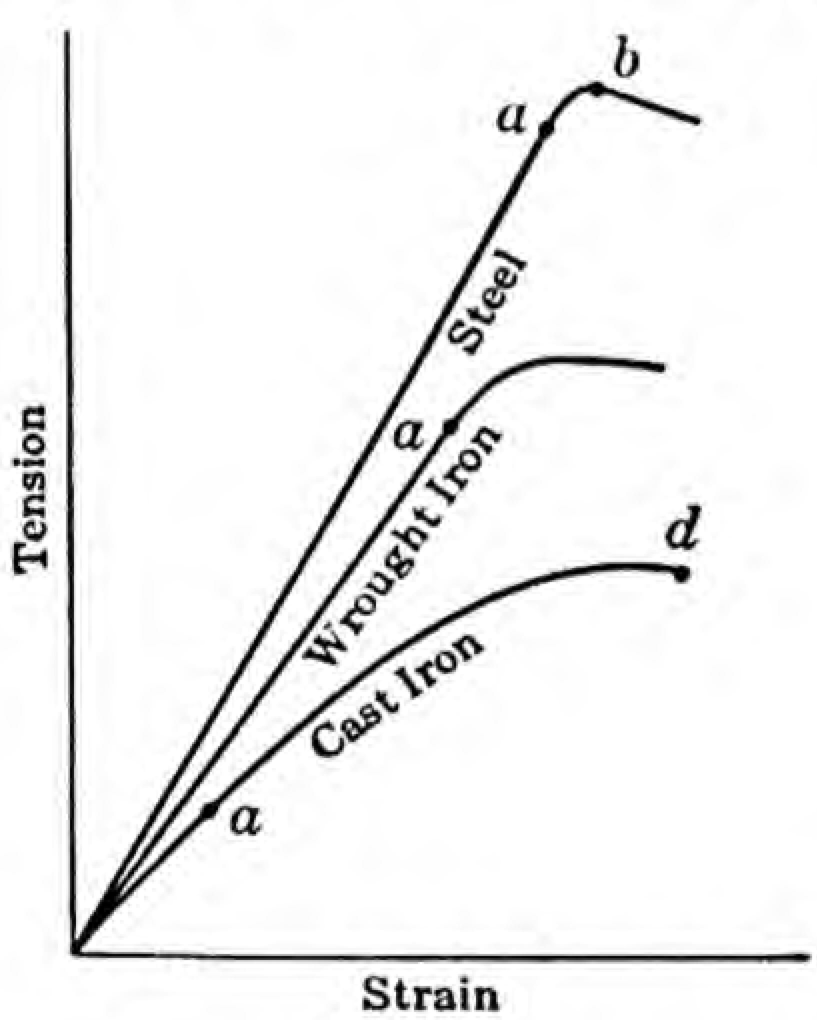


FIG. 21

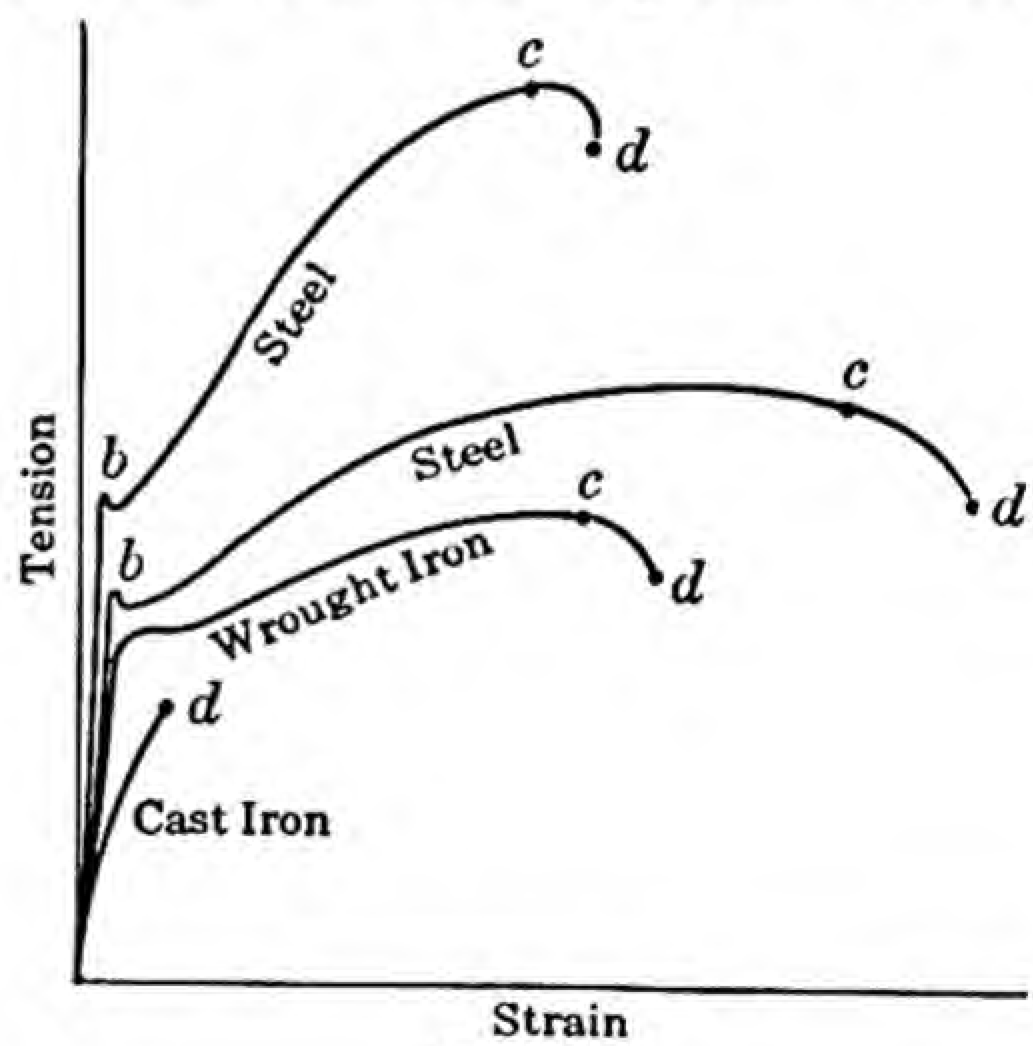


FIG. 22

case of wrought iron (Fig. 21), it is seen that there is a direct proportionality between the tension and the strain up to the point a. Near point a and for many substances practically coincident with it, is a point called the *elastic limit*. Beyond this point the material does not regain its former shape when the tension is removed. As the tension is increased beyond a, a point b, the yield point, is reached where the sample begins to lengthen considerably and the tension must be reduced in order to maintain any given value of the strain. After a certain stretch has occurred beyond the point b (Fig. 22), the material is able to withstand larger tensions until

the point c is reached. Then the material begins to thin down rapidly and at point d breaks in two.

The point a is rather indefinite. Its position depends on limits of accuracy specified. When one is able to distinguish a from the elastic limit, it is called the *proportional point.* In cast iron and softer metals there is scarcely any straight-line portion.

PROBLEMS

1. What must be the diameter of a drawn-steel wire 5 meters long if it is required to sustain a maximum load of 20 kilograms without stretching more than one millimeter?

SOLUTION: Substituting in Eq. 38 we have:

$$18.8 \times 10^{11} \frac{\text{dynes}}{\text{cm}^2.} = \frac{(20{,}000 \times 980)\text{ dynes} \times 500\text{ cm.}}{(\pi r^2)\text{cm}^2. \times 0.1\text{ cm.}}.$$

Hence, $r = 0.129$ cm. and $D = 0.258$ cm.

2. A is a drawn-steel wire and B is a copper wire. Compare the stretches of the two wires for the same load, A being twice as long as B and having twice the cross sectional area.

3. One end of a rod one meter long and 0.6 cm. in diameter is twisted through an angle of 0.2 radian. Where is the shearing strain the greatest? Where is it the least? Compute its maximum value.

4. The material of which a circular shaft is made will stand a maximum shearing strain of 0.002 without passing beyond the elastic limit. Through what angle can a shaft 25 ft. long and 1 in. in diameter be turned without this strain being exceeded? Explain clearly.

5. Show from Eq. 42 that for a horizontal bar of length l, clamped at one end and pulled down at the other end by a force F, the sag of the loaded end is given by the expression,

$$e = \frac{4\,Fl^3}{Ybd^3}.$$

6. For a certain material it is found that with a proper factor of safety a shaft may be safely twisted through no more than 1° for a length equal to 20 times the diameter of the cross section. What is the maximum strain in this case? Solve Problem 4 using this safety limit.

7. A long shaft is composed of two parts. One third of its length is aluminum. The rest is brass whose diameter is only half that of the aluminum portion. When one end of the shaft is twisted through a certain angle, what fraction of the twist occurs in each portion?

SLIDING FRICTION

25. Static Friction. — Let a smooth block of any material be laid on a smooth horizontal table (Fig. 23). Let a small force F, parallel to the table, be applied to the body. For a sufficiently small force, the body does not move. There is a reaction between

the surfaces of the block and the table. As the block is urged forward, it pushes forward on the table with a force f and the table pushes backwards on the block with an equal and opposite force, $-f$. The block is kept in equilibrium by the two forces, F and $-f$, which act equally in opposite directions upon it. As the force F increases, f increases and this continues until f cannot increase any further. Then F becomes slightly larger than f and the body starts into motion. The maximum value of f is called the maximum static friction force. *For smooth surfaces it is found that over quite wide ranges, the maximum static friction force* f_m *varies directly with the resultant force* (perpendicular to the surface) *which thrusts the two bodies together.* As long as F is parallel to the horizontal surface, the normal force N is equal to the weight of the body. However, if the applied force F acts at an angle to the horizontal surface, then the component of it which is normal to the surface must be added algebraically to the weight in order to give the resultant normal force N. We may write concerning the maximum friction force f_m,

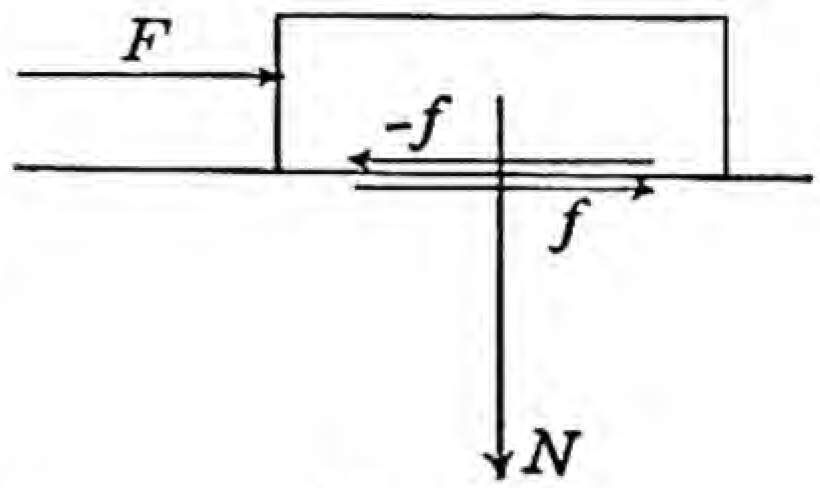


FIG. 23

$$f_m = \mu_s N. \tag{43}$$

The proportionality factor μ_s is called the coefficient of static friction. When it is multiplied by the normal force pushing two surfaces together, the product gives the maximum static friction force which acts parallel to the surface.

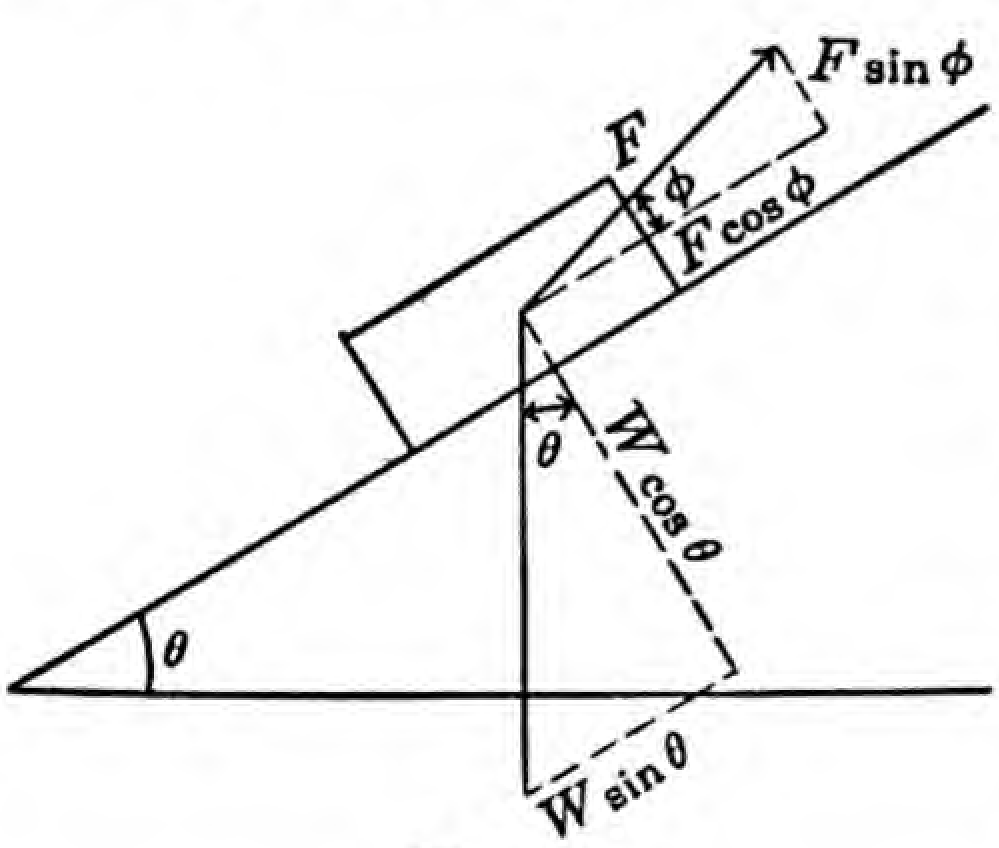


FIG. 24

The coefficient of friction is found to be independent of the area of contact and the size of N for very hard smooth surfaces which have no grain and which do not tend to seize or tear into each other.

As an example, consider a block of weight W resting on an incline of elevation θ (Fig. 24). A force F is applied making an angle ϕ with the plane. Suppose we wish to find how large F must be in order to pull the block up the plane. The normal

force between the surfaces is $W \cos \theta - F \sin \phi$. Hence the maximum static friction force is

$$\mu_s(W \cos \theta - F \sin \phi)$$

and since the friction force always opposes the motion, it is directed down the plane. The component $W \sin \theta$ also acts down the plane. In order that the block be started up the plane, $F \cos \phi$, the only force up the plane, must be ever so slightly larger than

$$\mu_s(W \cos \theta - F \sin \phi) + W \sin \theta.$$

The student should show that, if the force F is reversed (a thrust downward pushing the surfaces together as well as pushing the block down the incline),

$$F \cos \phi + W \sin \theta \geqq \mu_s(W \cos \theta + F \sin \phi)$$

is the condition for the starting of motion down the plane.

If the block is acted upon by no force except that of gravity, then if the angle of the incline is increased, the normal force $W \cos \theta$ gets smaller and hence the maximum friction force decreases. At the same time, the value of $W \sin \theta$, the component down the plane which tends to make the block slide, grows larger. The student should show that no matter what the weight of the body, there will be a given angle at which $W \sin \theta$ will become equal to the maximum friction force and the block will slide. Then

$$\mu_s = \tan \theta. \tag{44}$$

This critical angle is called the *limiting angle of repose* or the *angle of slip*.

26. Kinetic Friction. — After the necessary force F (Fig. 23) has been applied to a body to exceed just slightly the maximum static force, the body starts into motion with a very large acceleration. Since F has remained the same, we conclude that the friction force, after motion has started, has decreased considerably. Decreasing F, we find a value at which the body continues to move with a uniform velocity. Again the body is in equilibrium and that smaller value of F is numerically equal (though opposite) to the maximum kinetic friction force. This maximum kinetic friction force is likewise directly proportional to the normal force between the two surfaces.

$$f_m = \mu_k N. \tag{45}$$

The proportionality factor is called the coefficient of kinetic friction.

PROBLEMS

Solved Problem: Consider the case as shown in Fig. 24. Let W = 100 lbs., $\theta = 10°$, $\phi = 40°$, $\mu_s = 0.8$, and $\mu_k = 0.5$. Let F be reversed and equal 200 lbs. Take first the case where $\mu_s = 0.8$. The forces acting down the plane are 100 sin 10° and 200 cos 40°, giving a total of 160.6 lbs. The forces normal to the plane are 100 cos 10° and 200 sin 40°, or a total of 227 lbs. The friction force is the only force acting up the plane and its maximum value is $0.8 \times 227 = 181.6$ lbs. This is larger than 160.6 lbs., the force acting down the plane. Therefore the 100 lb. block will not start into motion and only 160.6 lbs. of friction force will be brought into play. If the block is given a start, then the kinetic friction force must be computed. Its maximum value is $0.5 \times 227 = 113.5$ lbs., which is less than 160.6 lbs. Therefore the block will slide with an acceleration

$$a = \frac{Fg}{W} = \frac{(160.6 - 113.5)\text{ lbs.} \times 32.2\,\frac{\text{ft.}}{\text{sec}^2.}}{100\text{ lbs.}} = 15.1\,\frac{\text{ft.}}{\text{sec}^2.}$$

1. A sled and contents weigh 500 lbs. The coefficient of static friction is 0.1 and of kinetic friction is 0.07. What pull on a rope parallel to the ground will start the sled? Keep it moving with a uniform velocity? Answer the same for the case of a person pushing the sled, the force being applied at a 30° angle with the ground.

2. Solve Problem 2 (a) § 12, when μ_k for M is 0.025.

3. In Problem 2 (b) § 12, what would be the largest coefficient of friction between m_1 and the plane down which it slides which will allow motion to continue uniformly when once started, the coefficient of friction between M and its plane being 0.025?

4. In Fig. 25, $\mu_s = 0.2$, $\mu_k = 0.1$. Will the system start into motion of its own accord? With what acceleration will the system move after the motion is started?

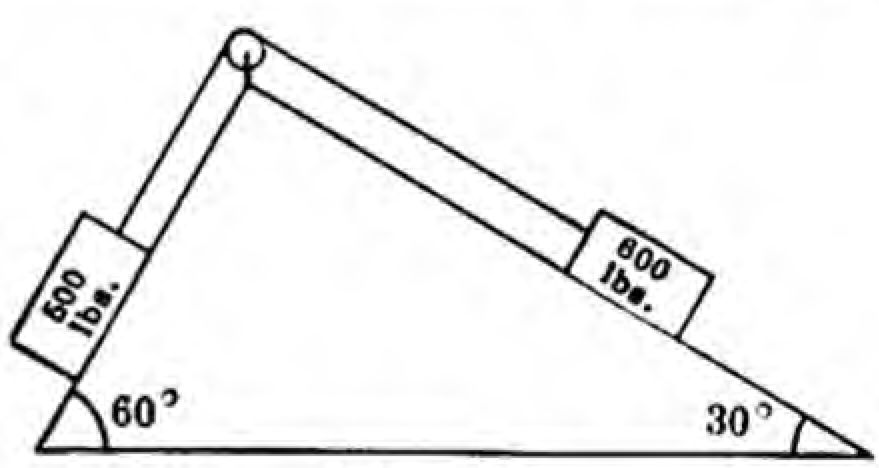


Fig. 25

5. A body weighing 50 lbs. rests on a horizontal surface. The coefficient of static friction is 0.25. A force is applied at an angle θ above the horizontal to push the body. Neglecting the weight of the body, compared to the size of the applied force, show that motion is impossible no matter how large the force is if

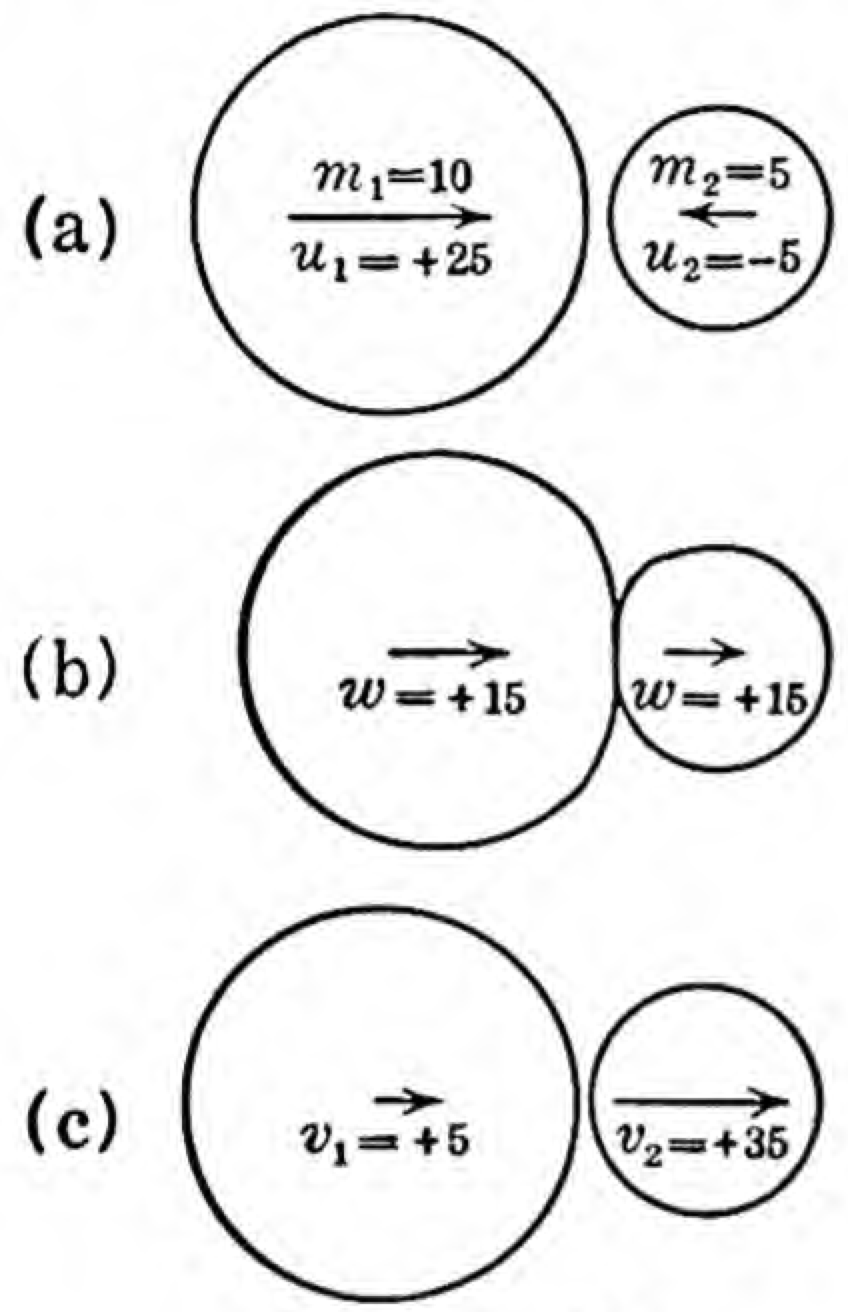
$$\tan\theta > \frac{1}{\mu_s}.$$

What force is required to start the motion when $\theta = 0°$, 10°, and 30°? What is the largest angle at which motion is possible?

6. The coefficient of static friction between rubber and a certain kind of pavement is 1.0. What is the maximum speed with which an automobile may go around a curve of radius 200 ft. if the pavement is level? At what angle must the curve be banked to have no side forces at this speed? In general show that if a curve is banked so that there is no sidewise force on a car which rounds the curve at a speed such that if the road were level it just would not slip, then the tangent of the angle of bank is equal to the coefficient of static friction.

Impact

27. Conservation of Momentum. — When two freely moving bodies collide, we have a very simple case to which we may apply Newton's laws of motion. Consider two spheres whose centers are moving along the same straight line (Figs. 26*a*, 27*a*). At the



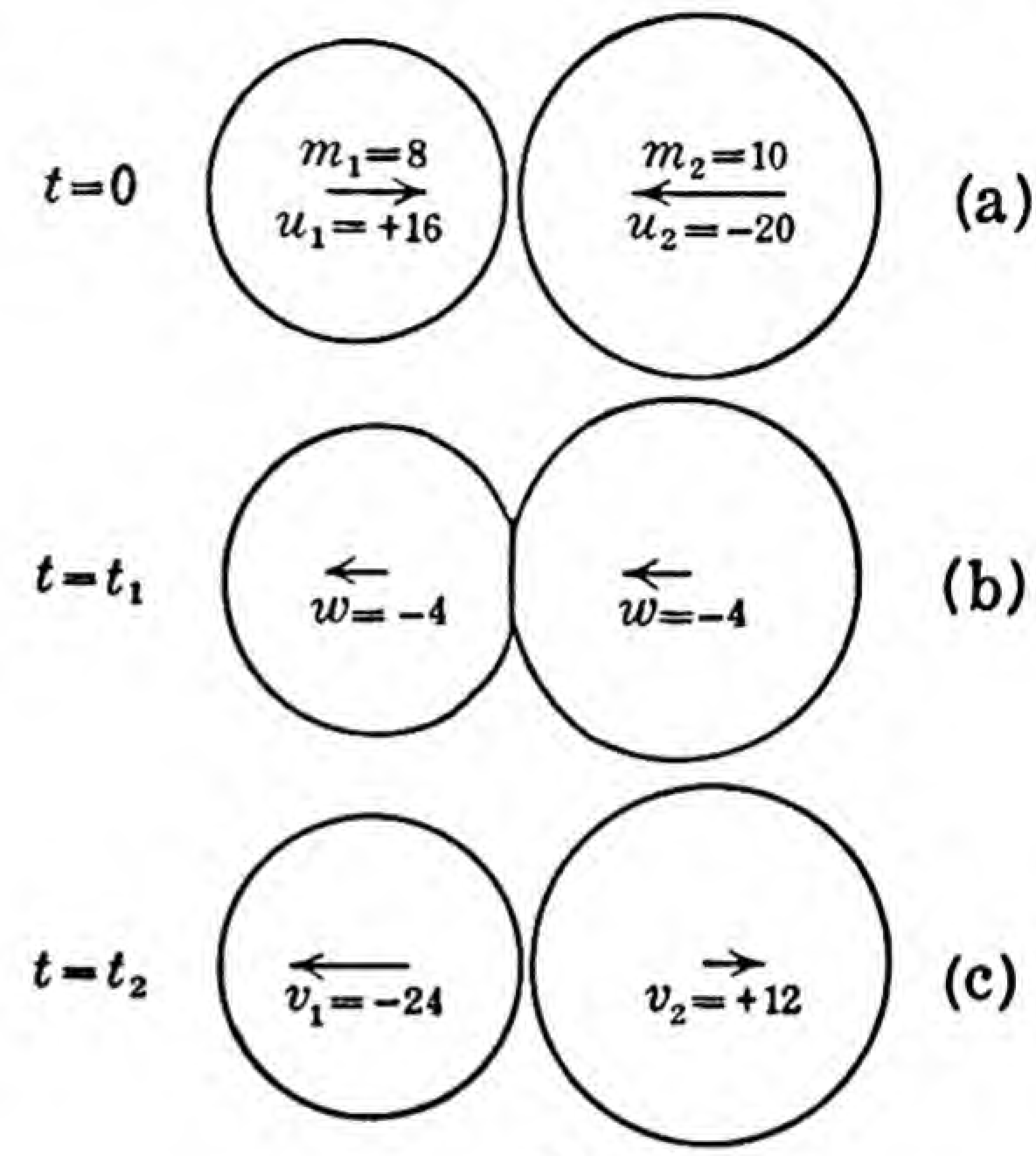


Fig. 26

Fig. 27

instant of touching no forces are acting on either body. During the collision, the bodies distort each other. The body A is pushed with a certain force F_1 to the *left* by body B. Simultaneously B is pushed to the *right* by body A, with a force F_2 exactly equal to F_1 but oppositely directed. These two equal forces increase from zero and reach a maximum when the bodies are most distorted (Figs. 26*b* and 27*b*) and then decrease to zero. This final state is reached in elastic bodies just when they are on the point of separating (Figs. 26*c* and 27*c*). Because the two forces are equal at every instant, then the impulse of the forces must be the same for both bodies over any time interval whatsoever. Let us call the time $t = 0$ at the first instant of impact and $t = t_2$ when the forces have decreased to zero. If we integrate over the whole time of collision, we obtain

$$\int_0^{t_2} F_1\,dt = -\int_0^{t_2} F_2\,dt = m_1(v_1 - u_1) = -m_2(v_2 - u_2),$$

or

$$m_1v_1 + m_2v_2 = m_1u_1 + m_2u_2. \tag{46}$$

This equation states a very important fact about impacts: *The vector sum of the momenta of two bodies after a collision is the same as the vector sum of the momenta of the bodies before the collision.* This is true no matter what the elastic properties of the bodies happen to be.

28. The Coefficient of Restitution. — Let the time $t = t_1$ be the time of maximum distortion of the colliding bodies, when they are relatively at rest and have a common velocity w. From Eq. 22 the impulse of compression is

$$\int_0^{t_1} F_1\,dt = -\int_0^{t_1} F_2\,dt = m_1(w - u_1) = -\,m_2(w - u_2) \quad (47)$$

and the impulse of restitution is

$$\int_{t_1}^{t_2} F_1\,dt = -\int_{t_1}^{t_2} F_2\,dt = m_1(v_1 - w) = -\,m_2(v_2 - w) \quad (48)$$

The ratio of the impulse of restitution to the impulse of compression is called the *coefficient of restitution* or coefficient of elasticity (not to be confused with modulus of elasticity). Calling this ratio e, we get from Eqs. 47 and 48,

$$e = \frac{v_1 - w}{w - u_1}, \text{ and } e = \frac{v_2 - w}{w - u_2}. \quad (49)$$

It is desirable to express e in terms of the initial and final velocities of the two bodies. The student should eliminate w from the two equations and obtain the following relation.

$$e = \frac{v_2 - v_1}{u_1 - u_2} = -\,\frac{v_1 - v_2}{u_1 - u_2}. \quad (50)$$

It is seen that $v_2 - v_1$ is the relative velocity of B with respect to A after the collision and $u_1 - u_2$ is their relative velocity before the collision. Note in Figs. 26 and 27 and from Eq. 50 that e is always a positive number. So we may write concerning these relative velocities,

$$\text{The coefficient of restitution} = \frac{\text{relative velocity after collision}}{\text{relative velocity before collision}}.$$

29. Perfectly Elastic and Perfectly Inelastic Collisions. — When $e = 1$, the collision is said to be perfectly elastic. From Eq. 50 we see that the relative velocity of the bodies after collision is equal and opposite to their relative velocity before collision. The values given in Figs. 26 and 27 have been computed for $e = 1$.

When $e = 0$, $v_2 - v_1 = 0$, $v_1 = v_2$, and the bodies do not separate. Such a collision is said to be perfectly inelastic.

30. Energy Relations. — When a collision is inelastic and the relative velocity of the colliding bodies is decreased, we might expect that the system of bodies has lost kinetic energy. To investigate this, we shall obtain an expression for the difference between the kinetic energy of the two bodies before and after collision. Call this difference in energy ΔE.

$$\begin{aligned}\Delta E &= \tfrac{1}{2}\, m_1u_1^2 + \tfrac{1}{2}\, m_2u_2^2 - \tfrac{1}{2}\, m_1v_1^2 - \tfrac{1}{2}\, m_2v_2^2\\ &= \tfrac{1}{2}\, m_1(u_1^2 - v_1^2) + \tfrac{1}{2}\, m_2(u_2^2 - v_2^2)\\ &= \tfrac{1}{2}\, m_1(u_1 - v_1)(u_1 + v_1) + \tfrac{1}{2}\, m_2(u_2 - v_2)(u_2 + v_2). \qquad (51)\end{aligned}$$

Besides this relation between the four velocities, we have Eqs. 46 and 50:

$$m_1v_1 + m_2v_2 = m_1u_1 + m_2u_2, \qquad (46)$$

$$v_2 - v_1 = e(u_1 - u_2). \qquad (50)$$

We thus have three equations and four variables, u_1, u_2, v_1, and v_2, so we may therefore combine the equations so as to reduce them to one equation containing two variables. It would be effective to express ΔE as a function of either the initial velocities u_1 and u_2 or of the final velocities v_1 and v_2. Let us choose the former. We might proceed in the standard way to eliminate v_1 from Eqs. 46 and 50 and likewise from Eqs. 50 and 51 and then eliminate v_2 from these two resulting equations, but the following simpler though less obvious method will be used.

Substituting the value of $m_1(u_1 - v_1)$ from Eq. 46 into Eq. 51,

$$\Delta E = \tfrac{1}{2}[m_2(v_2 - u_2)(u_1 + v_1) + m_2(u_2 - v_2)(u_2 + v_2)].$$

$$\Delta E = \tfrac{1}{2}\, m_2(v_2 - u_2)(u_1 + v_1 - u_2 - v_2), \qquad (52)$$

which from Eq. 50 becomes

$$\Delta E = \tfrac{1}{2}\, m_2(v_2 - u_2)[u_1 - u_2 - e(u_1 - u_2)],$$

$$\Delta E = \tfrac{1}{2}\, m_2(v_2 - u_2)(u_1 - u_2)(1 - e).$$

We must now obtain $(v_2 - u_2)$ in terms of u_1 and u_2. Substituting v_1 from Eq. 50 into Eq. 46,

$$m_2v_2 - m_2u_2 + m_1v_2 = m_1u_1 + m_1eu_1 - m_1eu_2.$$

Subtracting m_1u_2 from each side,

$$(m_1 + m_2)(v_2 - u_2) = m_1(1 + e)(u_1 - u_2).$$

So
$$\Delta E = \frac{1}{2}\frac{m_1m_2}{m_1 + m_2}(1 - e^2)(u_1 - u_2)^2. \qquad (53)$$

If the collision were perfectly elastic, $e = 1$ and $\Delta E = 0$. This fact could have been got directly by using $e = 1$ in Eq. 50 and

then Eq. 52 would have given directly that $\Delta E = 0$. There is no loss in kinetic energy in a perfectly elastic collision. All the potential energy stored up in the deformed bodies is changed back into kinetic energy when the bodies regain their shapes completely before separating. In inelastic collisions where there is internal friction between the molecules of the body, some of the energy appears as heat. A small part of the energy leaves the system as sound waves in the air. In an electric arc and other places, an electron may collide with an atom in an inelastic manner so that energy is dissipated as light waves. It is to be noted that no matter how elastic or inelastic the bodies are or no matter how much energy may be changed into heat energy causing the relative velocity of the bodies after collision to be ever so small, still Eq. 46 holds and the momentum is always completely conserved.

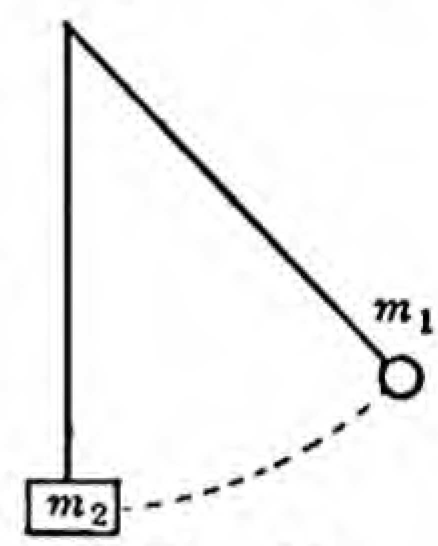


Fig. 28

Consider the standard laboratory experiment shown in Fig. 28. Two pendulums of the same length are supported on the same axis. The mass m_1 is raised to a certain position and released. If some clay, wax, or a mechanical device causes the two bodies to stick together, then we have the case where $e = 0$ and since m_2 is initially at rest, $u_2 = 0$. In this special case

$$\Delta E = \frac{m_2}{m_1 + m_2} \left(\tfrac{1}{2}\, m_1 u_1^2\right).$$

That is, a definite fraction of the original total energy of the system is lost. That fraction is the ratio of stationary mass to the total mass and is independent of the velocity of the falling mass. This equation just derived may be proved experimentally; the kinetic energies are determined by the vertical heights through which m_1 falls and the height through which the combined mass rises on the opposite side.

31. Glancing Collisions. — The theorems above have been derived for collisions in a straight line. If we consider two spheres moving so as to collide at any angle whatsoever, it may be shown that, if we consider collisions such that rotation of the spheres does not take place, Eq. 46 still holds as a vector summation. This is seen directly from the equations from which Eq. 46 is derived. If rotation is excluded so that no change of angular momentum may occur, then the only change in momentum which the impulse can

produce is $m_1(v_1 - u_1)$. Thus in Fig. 29 the original momentum vectors m_1u_1 and m_2u_2 are shown. If the two smooth spheres collide so that AB is the line through their centers, the forces of restitution will act along AB. The velocities after collision are related so that the impulse vectors OA and OB (equal and opposite) are respectively equal to $m_1(v_1 - u_1)$ and $m_2(v_2 - u_2)$ as seen in the figure.

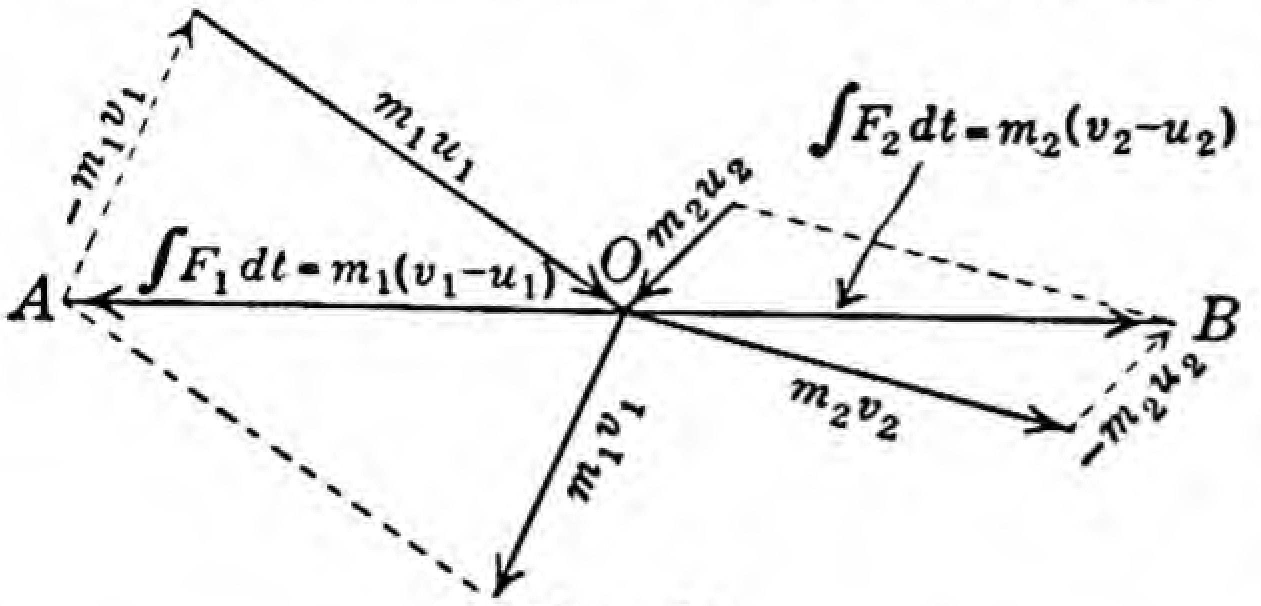


FIG. 29

PROBLEMS

SOLVED PROBLEM: Let us solve the case illustrated in Fig. 26 for the case where $e = 0.3$. From Eq. 46 we obtain

$$10\,v_1 + 5\,v_2 = 10 \times 25 - 5 \times 5.$$

From Eq. 50 we obtain $0.3\,(25 + 5) = v_2 - v_1$.
Solving these two equations simultaneously we get $v_1 = +12$, $v_2 = +21$. The units of v_1 and v_2 must be the same as those of u_1 and u_2. The plus signs indicate that after the collision both bodies are traveling to the right. For the case shown in the figure, e was taken as unity. It is to be noted that as e becomes less than unity, the velocities become more nearly equal. In the case just worked out the difference in v_1 and v_2 is only 9 while in the figure it is 30.

1. In Figs. 26 and 27, the collisions have been considered perfectly elastic. Find w, v_1, and v_2 and check the equalities of the vector differences predicted in Eqs. 47 and 48.

2. Solve for v_1, v_2, and w for the following cases:

m_1	m_2	u_1	u_2	e	ANSWERS		
					v_1	v_2	w
8	10	+ 16	− 20	0.3	− 10.0	+ 0.8	− 4
8	10	+ 16	− 20	0.2	− 8.0	− 0.8	− 4
8	10	+ 16	− 20	0.1	− 6.0	− 2.4	− 4
8	10	+ 16	− 20	0.0	− 4.0	− 4.0	− 4

Note the gradual change from the condition shown in Fig. 27 as the perfectly inelastic impact is reached. That w depends upon the initial conditions and not on e is seen from Eq. 47. Calculate the energy lost in each collision.

3. Write down the momentum equation for two bodies separated by an explosion or by a released coiled spring. Show that the heavier the gun, the greater is the velocity of the bullet and the greater is the fraction of the total energy of explosion that is imparted to the bullet. Neglecting losses, if the gun has an exceedingly large mass compared to the bullet, the bullet gets practically all the energy of the explosion.

4. If a set of highly elastic steel balls are placed in contact in a smooth groove and any number of balls, n say, in contact, are rolled against the set with a given velocity, the balls will come to rest in contact with the others and the same number of balls will leave the other end with the same velocity. Prove that this is so by the impact equations for perfect elasticity and show why it is impossible for twice the number of balls to leave with half the velocity of the colliding ones.

5. Prove that when two perfectly elastic bodies of the same mass moving in the same straight line collide, their velocities interchange.

6. Show from Eq. 50 that when a body of small mass collides with a stationary body of exceedingly great mass, for perfect elasticity, the velocity of rebound is nearly equal to the velocity of impact, while for inelastic impacts, the coefficient of restitution is nearly the ratio of the velocity after impact to that before impact.

7. Show that if a ball is dropped from a height h onto a heavy plate whose mass is very large compared to that of the ball, and bounces to a height h', the value of e for those two substances is practically,

$$e = \sqrt{\frac{h'}{h}}.$$

8. When two smooth spheres whose masses and momenta are given collide at an angle different from 180°, show by vector diagrams that if the direction of rebound of both spheres is known, then the velocities after rebound can be determined as well as the impulse of the collision.

9. Show in perfectly elastic collisions along a straight line that the common velocity had by two bodies at maximum distortion is the algebraic mean of the initial and the final velocities of either body.

Angular Motion

32. Angular Units. — Consider a circle of radius r (Fig. 30). An angle such as AOB which subtends an arc equal to the radius is defined to be one *radian*. From this it follows that when an angle θ (DOC) is expressed in radians, the length of arc s (arc DC) subtended by this angle is given by

$$s = r\theta. \tag{54}$$

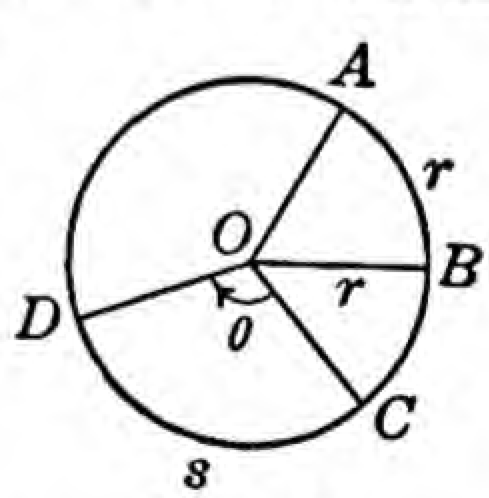


Fig. 30

In this equation s will have the same units as r, and θ is a pure number.

If the radius OC rotates about O, the rate of change of the angle θ is called the angular velocity of the line. This physical quantity is usually indicated by ω. Its units are radians per second. From Eq. 54, where v is the velocity of the point C,

$$v = \frac{ds}{dt} = r\frac{d\theta}{dt} = r\omega. \tag{55}$$

If the angular velocity is not constant, then the relation between the rate of change of ω, or α, and the rate of change of the velocity of point C is found by differentiating Eq. 55.

$$a = \frac{dv}{dt} = r\frac{d\omega}{dt} = r\alpha. \tag{56}$$

From these last three equations we see that the angular quantity when multiplied by the radius gives the corresponding linear quantity along the circumference.

33. Uniformly Accelerated Angular Motion. — For this type of motion, α is constant. From the above paragraph,

$$\theta = \int \omega\, dt, \tag{57}$$

and

$$\omega = \int \alpha\, dt. \tag{58}$$

Since α is a constant, Eq. 58 gives $\omega = \alpha t + c$.

As in § 4, we may select an initial condition such as $\omega = \omega_0$ when $t = 0$ and

$$\omega = \alpha t + \omega_0. \tag{59}$$

Substituting this in Eq. 57,

$$\theta = \int (\alpha t + \omega_0)dt = \frac{\alpha t^2}{2} + \omega_0 t + c'.$$

Suppose that we choose to measure the angular displacement from the position of the body when we started to count time. Then $\theta = 0$ when $t = 0$, and

$$\theta = \tfrac{1}{2}\alpha t^2 + \omega_0 t. \tag{60}$$

Compare Eqs. 59 and 60 with Eqs. 5 and 8 in § 4.

PROBLEMS

1. Show that an angle is a dimensionless quantity.

2. Eliminate t from Eqs. 59 and 60 and obtain the equation similar to Eq. 9:

$$\omega^2 = \omega_0^2 + 2\,\alpha\,\theta. \tag{61}$$

Also obtain

$$\theta = \frac{\omega + \omega_0}{2}\,t.$$

3. At the instant power is cut off from a motor, the armature is making 1750 r.p.m. A constant torque brings it to rest in 10 seconds. Find (*a*) acceleration; (*b*) total number of revolutions before stopping; (*c*) the distance it would have traveled if it had been rolling on the ground and had been brought to rest during the above time. Its diameter is 20 cm.

4. Find expressions for ω and θ for motion in which $\alpha = 0$.

5. A cylinder of 1 foot radius rolls down an inclined plane 100 feet long in 8 seconds. Calculate the angular acceleration and the angular velocity at the bottom of the plane.

6. A fly wheel revolving at 5 rev. p. sec. is brought to rest in 10 revolutions. What is the angular acceleration? How long is required to stop it?

Center of Mass

34. Center of Mass. — *The center of mass of a body is defined as that point, located usually within the body, at which a single force, applied in any direction, will give the body a motion of translation only, i.e.*, there will be no rotation and all particles of the body will move with the same acceleration.

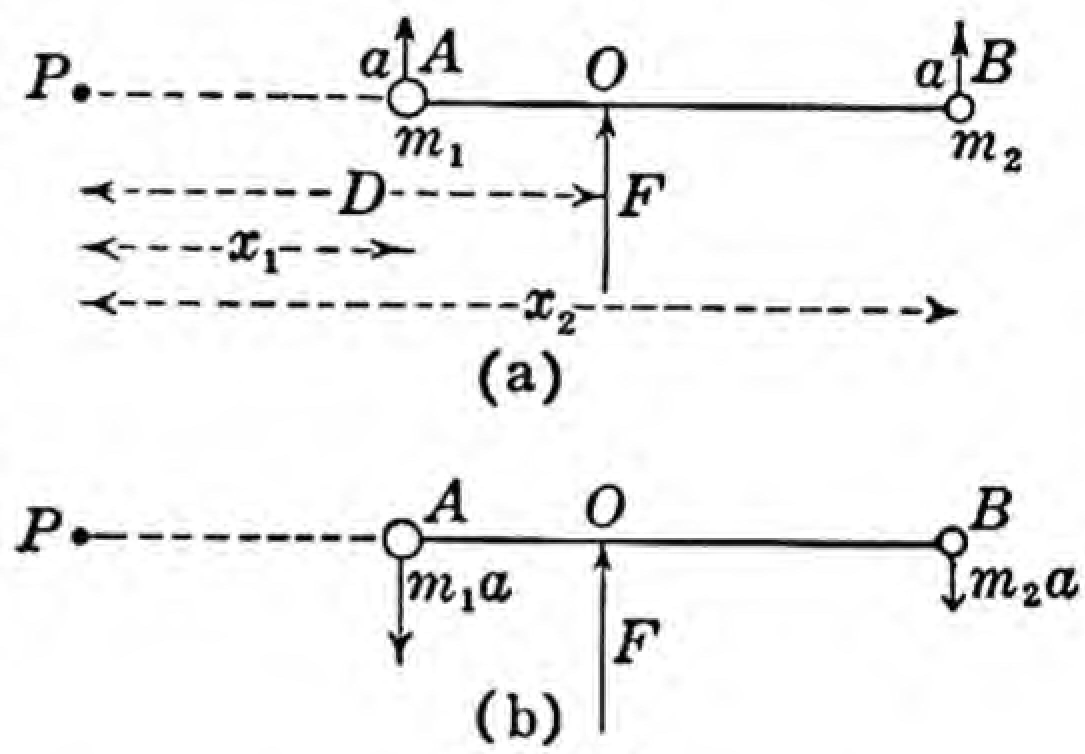


Fig. 31

Let us consider the ideal simple case of two particles, m_1 and m_2, held rigidly together by a massless connection as shown in Fig. 31*a*.

We wish to find the location of the point O such that the force F, applied at O perpendicular to AB, will give both masses m_1 and m_2 an acceleration of

$$a = \frac{F}{m_1 + m_2},$$

parallel to F. We shall first have to examine all the forces involved. By means of the connection, a force m_1a must have been applied to m_1 and a force m_2a applied to m_2. Then we know that these two masses must react and push backward on the connecting system with the corresponding equal and opposite forces. Since the system as a whole must have no rotational acceleration, neither will the connecting system. The connecting system is acted upon by three forces (Fig. 31*b*), F at O, m_1a at A, and m_2a at B. The system is in rotational equilibrium under the action of these forces, so for a point P, distant D from O, we may write (§ 14):

$$(m_1 + m_2)aD - m_1ax_1 - m_2ax_2 = 0.$$

Hence
$$D = \frac{m_1x_1 + m_2x_2}{m_1 + m_2}. \tag{62}$$

For three masses on a straight line

$$D = \frac{m_1x_1 + m_2x_2 + m_3x_3}{m_1 + m_2 + m_3}, \text{ etc.}$$

PROBLEMS

1. Prove Eq. 62 for the case when F makes an angle different from 90° with the line AB.

2. Show that the center of mass must always be somewhere on the line AB.

35. The Center of Mass of a Rigid Body. — An extended rigid body of mass M will be referred to a rectangular system of co-ordinates and the x, y, and z co-ordinates of the center of mass will now be obtained.

In order to involve only the x co-ordinates of each particle in the body, let force F lying in a plane perpendicular to the x axis be applied at the center of mass O (Fig. 32), whose co-ordinates are $(\bar{x}, \bar{y}, \bar{z})$. By definition, the position of O is such that the whole body will move in the direction of the force F with a linear acceleration

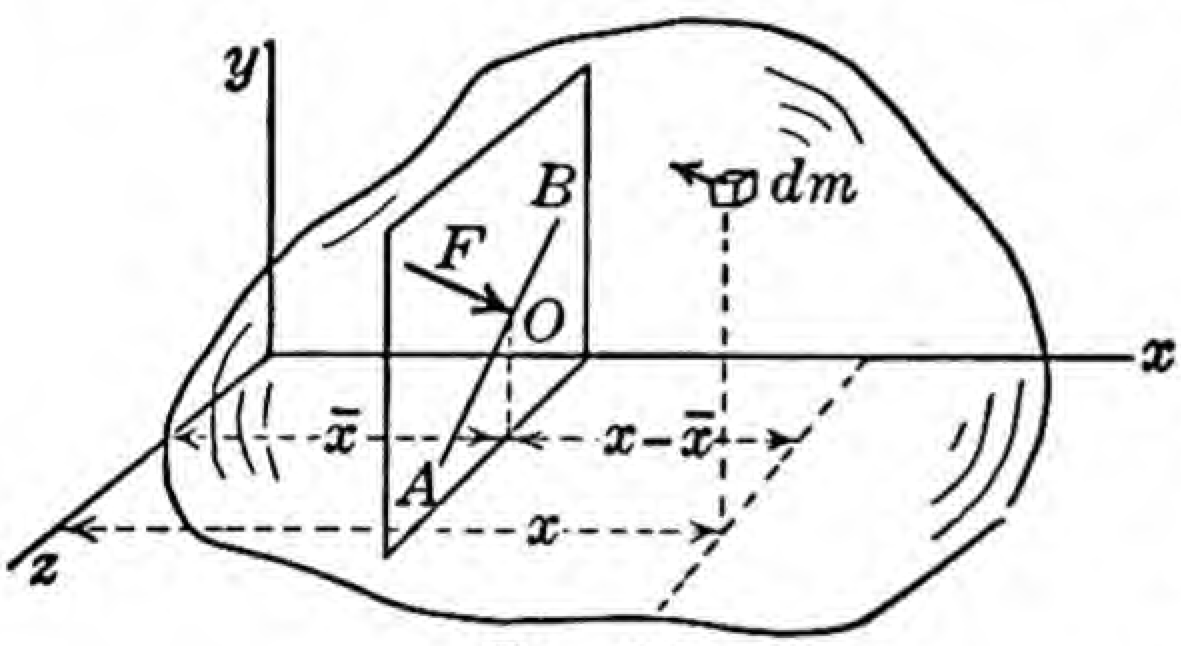


FIG. 32

$$a = \frac{F}{M}.$$

Each atom of the body is attached to its neighboring atoms by a system of electric forces which maintain the shape of the body. When the atoms about O are moved slightly out of their normal equilibrium position, they in turn pull on neighboring atoms. Thus the body is eventually slightly distorted, by such an amount that each elementary mass dm is pushed forward with a force necessary to give it the common acceleration a. In a rigid body, the massless connecting system referred to in the previous paragraph is the elastic medium through which act the electric forces between the atoms. If the connecting system pushes forward on each mass dm of the body with a force equal to $a \cdot dm$, then by Newton's third law, each mass reacts backward on the connecting system with an equal and opposite force, as shown in the figure. Thus the connecting system has one force F, and a series of parallel forces in the opposite direction scattered throughout the whole solid body. Consider an axis AB through O perpendicular to both F and the x axis. The lever arm of the reaction force $a \cdot dm$ about the axis AB is the distance $x - \bar{x}$. By definition, the point O is situated far enough out along the x axis that the moments about AB of all the forces on the connecting system to

the right of O (for which $x - x$ is positive) must be counterbalanced by those on the left of O (for which $x - \bar{x}$ is negative). Using the second condition of equilibrium, we may write for the axis AB,

$$\Sigma[a\, dm(x - \bar{x})] = 0.$$

Therefore
$$\int x\, dm = \int \bar{x}\, dm = \bar{x} \int dm.$$

So the x co-ordinate of the center of mass is

$$\bar{x} = \frac{\int x\, dm}{\int dm}.$$

Similarly
$$\bar{y} = \frac{\int y\, dm}{\int dm}, \tag{63}$$

and
$$\bar{z} = \frac{\int z\, dm}{\int dm}.$$

36. Velocity and Acceleration of the Center of Mass.—The value of $\bar{x}$ from Eq. 63 may be written in the form

$$\bar{x} = \frac{x_1\, dm_1 + x_2\, dm_2 + \cdots}{dm_1 + dm_2 + \cdots}. \tag{64}$$

If the rigid body is in motion, we may obtain the x component of the velocity of the center of mass V_x by differentiating this expression with respect to time, term by term, obtaining

$$\frac{d\bar{x}}{dt} = V_x = \frac{\frac{dx_1}{dt}\, dm_1 + \frac{dx_2}{dt}\, dm_2 + \cdots}{dm_1 + dm_2 + \cdots} = \frac{\int v_x\, dm}{\int dm}, \tag{65}$$

and again

$$\frac{d^2\bar{x}}{dt^2} = A_x = \frac{\frac{d^2x_1}{dt^2}\, dm_1 + \frac{d^2x_2}{dt^2}\, dm_2 + \cdots}{dm_1 + dm_2 + \cdots} = \frac{\int a_x\, dm}{\int dm}. \tag{66}$$

Eqs. 65 and 66 give us the x components of the velocity and acceleration of the center of mass in terms of the x components of the velocities and accelerations of the individual particles which make up the mass.

Each term of the numerator in Eq. 66 is the x component of the force on each dm, which gives it its acceleration along the x axis.

By following through the connecting system and applying Newton's second and third laws of motion, we find that $\int a_x\,dm$ equals the resultant of all the x components of the forces applied externally to the body. So we may write

$$A_x = \frac{\Sigma F_x}{M}. \tag{67}$$

Likewise for the y and z components and the total force. Therefore

$$A = \frac{\Sigma F}{M}. \tag{68}$$

Hence it may be stated that the acceleration of the center of mass of a rigid body is equal to the quotient of the vector sum of all externally applied forces by the total mass of the body. If the vector sum of all the externally applied forces is zero, the center of mass of the body can have no acceleration. Only the external forces need be considered because any forces internal to a body must have their reactions on neighboring portions of the same body and the center of mass cannot be moved by them. Thus, if an explosion occurs in a stationary body, the separate parts of the mass are at every instant so situated that the center of mass of the whole set of parts is the same as that for the original single body.

37. Action of a Couple. — A couple is composed of two equal and oppositely directed, parallel, but not concurrent forces. If a couple acts upon a body, the vector sum of the applied force in any direction is zero; it therefore follows (Eq. 68) that the acceleration of the center of mass is zero. Therefore, we must conclude that the rotation produced by a couple, no matter where applied, is always about an axis through the center of mass. As an example consider a rod at rest (Fig. 33), with center of mass at O and let the same couple be applied in the different positions as shown. Applying Eq. 68, we find that the acceleration of the center of mass is zero. Since there is an applied torque, the body must rotate, but since the center of mass must remain at rest, rotation must take place about an axis through it. If the couple is simply transferred parallel to itself, exactly the same effect will be pro-

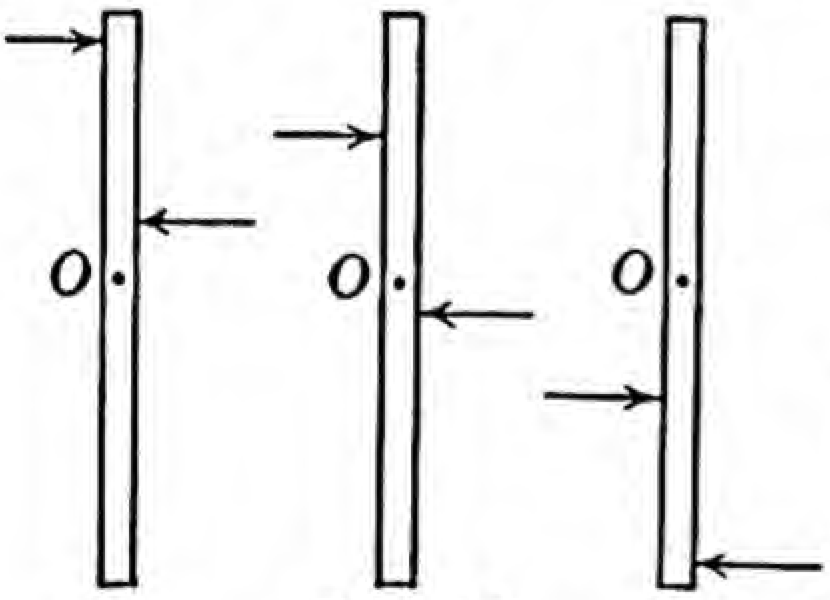


FIG. 33

duced in each case, *i.e.* exactly the same rotation about the same axis through O. If the plane in which the couple acts is changed, then the rotation will be about a different axis but one which must pass through O.

38. Motion of a Rigid Body when the Resultant of the Forces Acting upon It Is Not Zero and Does Not Pass through the Center of Mass.—Let F (Fig. 34) be the resultant of the forces applied to a body whose center of mass is at O. Without affecting the motion, we may apply at O equal and opposite forces, F_1 and F_2, each equal in magnitude to F and parallel to it. By so doing we are enabled to resolve the motion into one of pure rotation about the center of mass, O, produced by the couple F and F_2, and one of pure translation produced by the force F_1 acting through the center of mass.

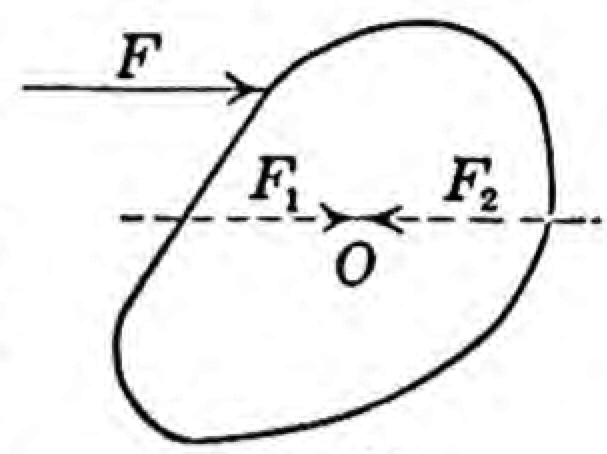


FIG. 34

We may state, therefore, that regardless of where a force is applied to a rigid body, the center of mass of the body will have the same acceleration as if a parallel and equal force acted through its center of mass. Furthermore, using the results of the previous paragraph, we may state that the rotation produced by the force takes place about an axis perpendicular to the force and through the center of mass, the moment of the force being the product of the force and its perpendicular distance from the axis.

39. Center of Gravity.—Consider a body of mass M (Fig. 35a) and let a force F act through its center of mass. As we have seen

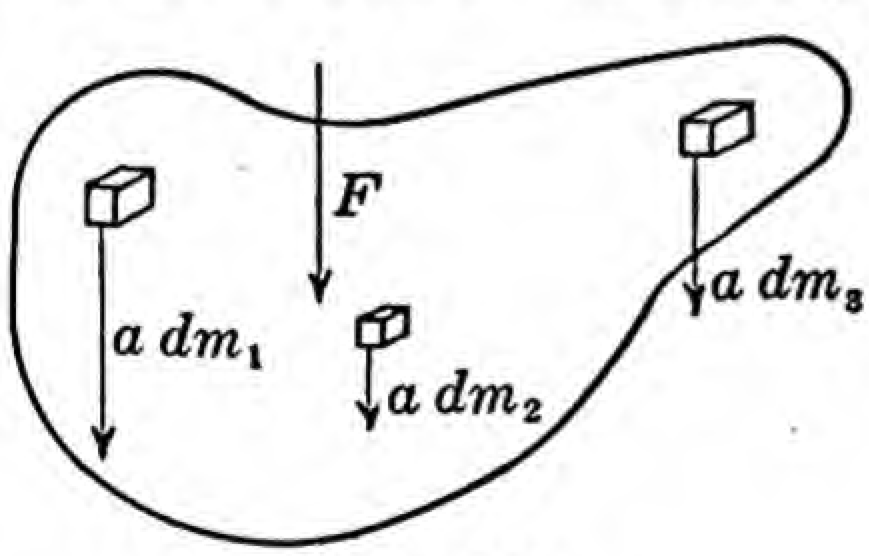


FIG. 35a

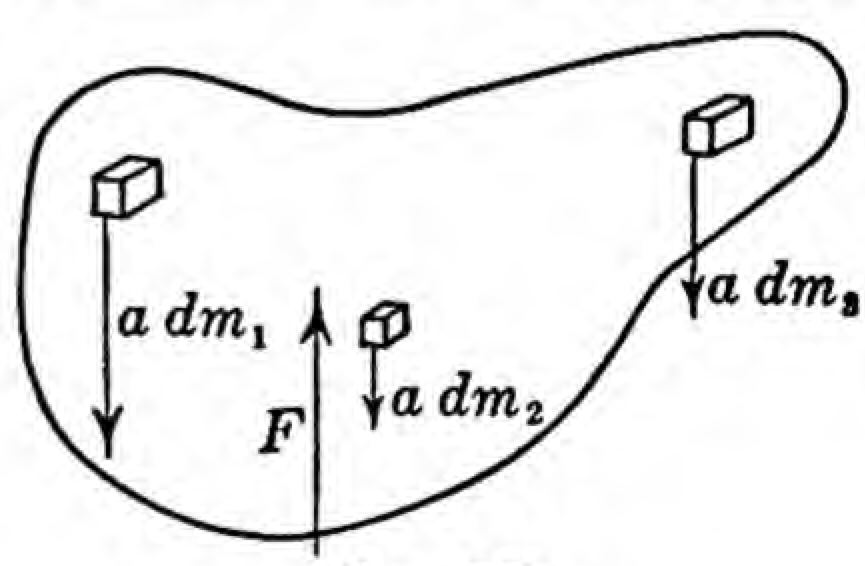


FIG. 35b

in §§ 34 and 35, the connecting system distributes the force so that each mass is given the same acceleration and there is no rotation. The mass dm_1 is acted upon by a force $a\ dm_1$, the mass dm_2 by a force $a\ dm_2$, etc. Thus each force is proportional to the mass of the particle on which it acts. Furthermore, the forces are all parallel and their sum is equal to F.

Now suppose that F be removed and by some external means those same forces be applied to the individual masses. Then if a force F, reversed in direction, be applied through the center of mass (Fig. 35*b*) the body will be in equilibrium under an equal and opposite set of parallel forces.

Let us next consider the gravitational attraction on this same mass. Each mass dm is acted upon by a gravitational force $g\,dm$, the sum of all these forces making up a total of W, the weight of the body. Just as in the above case, the force on each mass is proportional to the mass of that particle. Therefore we see that this case is the same as outlined above and so by applying a force W in place of F in Fig. 35*b*, the body must be in equilibrium.

If the gravitational forces are all parallel and if g is constant over the whole body, then we see from the above argument that a single upward force W acting through the center of mass will keep the body in equilibrium. However, there are three ways in which these conditions are not precisely met. (1) The gravitational forces on different portions of a body are not parallel but are convergent, since the forces on the various portions of the body are all directed to the center of the earth. (2) The gravitational force on a mass varies inversely as the square of its distance from the center of the earth, so the particles on the lower side of a body are acted upon with greater forces in proportion to their masses than the particles at the top of a body. There are balances which can demonstrate that two weights when side by side weigh more than when one is set on top of the other. (3) As mentioned in § 11, the value of g varies in a horizontal direction, so in an extended mass the forces on parts of a body may be greater than the forces on equal masses in other parts of the body in the same horizontal level. The second of these items shows us that the effective center of the pull of gravity is always very slightly below the center of mass. The point where a single force may be applied which will balance the gravitational forces is called the center of gravity of a body. It is in only the most precise work that any distinction must be made between the center of mass and the center of gravity.

It is easily seen from Fig. 35 that in order to simplify problems we may replace the gravitational forces acting on a rigid body of weight W by a single force W applied downward at the center of gravity of the body.

PROBLEMS

1. Find the center of mass of an isosceles triangle of height h and base b. In Fig. 36 the equation of the line forming one side of the triangle is

$$y = \frac{1}{2}\frac{b}{h}x.$$

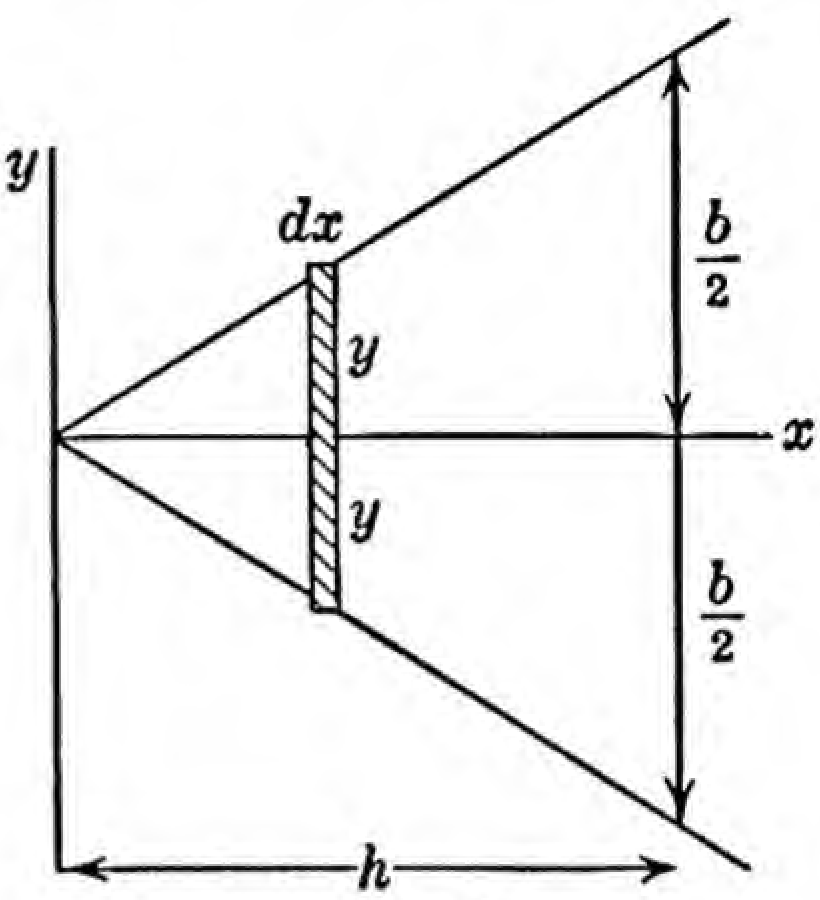


FIG. 36

All the shaded area of width dx has the same value of x and may be used as a lump in the expression dm. If ρ is the mass of each unit of area, then,

$$dm = \rho\, 2\, y\, dx.$$

So

$$\bar{x} = \frac{\int_0^h \rho\, 2\, yx\, dx}{\int_0^h \rho\, 2\, y\, dx} = \frac{\int_0^h \rho \frac{b}{h} x^2 dx}{\int_0^h \rho \frac{b}{h} x\, dx} = \frac{\left.\frac{x^3}{3}\right]_0^h}{\left.\frac{x^2}{2}\right]_0^h} = \frac{2}{3}h.$$

It is seen that no matter what the base of an isosceles triangle may be, the center of mass is at a point two thirds of the distance from the apex to the midpoint of the base.

2. Find the center of mass of the triangle in Problem 1 when the density varies directly as the distance from the apex. Let

$$\rho = cx.$$

Then

$$\bar{x} = \frac{\int_0^h cx \frac{b}{h} x^2\, dx}{\int_0^h cx \frac{b}{h} x\, dx} = \frac{\int_0^h x^3\, dx}{\int_0^h x^2\, dx} = \frac{3\, h}{4}.$$

As was to have been expected, the center of gravity is nearer the base than in the above problem.

3. Find the center of mass of a right cone of height h, density ρ, and radius of base r.

4. Solve Problem 3 for a density varying directly as the distance vertically down from the apex.

5. Solve Problem 3 for the case where the density is 8 gms./cc. for $0 < x < h/2$ and is 12 gms./cc. for $h/2 < x < h$.

6. The density of air at the earth's surface is 0.00128 gms./cc. Three hundred meters above the surface it is 0.00120 gms./cc. Find the center of mass of a column of air 1 sq. meter in cross section and 300 meters high. Assume that in this range the density is a linear function of the height.

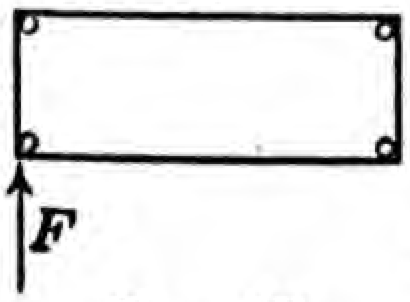


FIG. 37

7. A rectangular board (Fig. 37) weighing 10 kg. rests on universal roller bearings on a horizontal smooth surface (*i.e.* it is free to move without friction in any direction). A force, F, of 300 gms. is applied to the body for 0.1 sec. In what direction will the center of mass of the body move, and what will be the velocity of its center of mass? What other motion will it have?

8. A bar of mass m (Fig. 38) is dropped from an aeroplane traveling northward with velocity v_0. It receives a horizontal blow, F (in the nature of a kick), southward just as it leaves the machine. Discuss the motion as to nature and direction. Discuss the motion if the impulse of F (1) equals mv_0, (2) is greater than mv_0, (3) is less than mv_0.

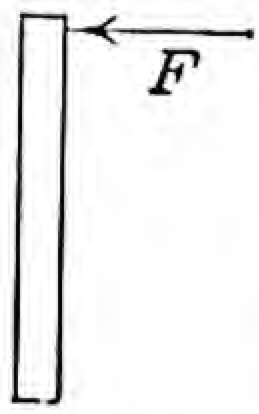


FIG. 38

9. Find the center of mass of the hollow cone formed by removing from the inside of a right circular cone a cone having the same axis and vertex but with a circular base of smaller radius.

Moment of Inertia

40. Fundamental Concepts. — In order to increase the angular velocity of a body about a given axis, a moment of force is required. Experiment shows that for a rigid body the angular acceleration is directly proportional to the moment of the force applied, *i.e.*

$$L = I\alpha. \tag{69}$$

This is a special statement of Newton's second law of motion. The factor I is called the moment of inertia of the body about the particular axis considered. It is the inertia factor corresponding to mass in linear motion. However, as we shall find later, in rotary motion the inertia factor depends not only upon the mass of the body but also upon the distribution of that mass about the axis of rotation. We shall proceed to show that

$$I = \int r^2\, dm,$$

where dm is any infinitesimal mass and r is its distance from the axis rotation. This physical quantity is sometimes called the second moment of a body since the distance enters as the second power.

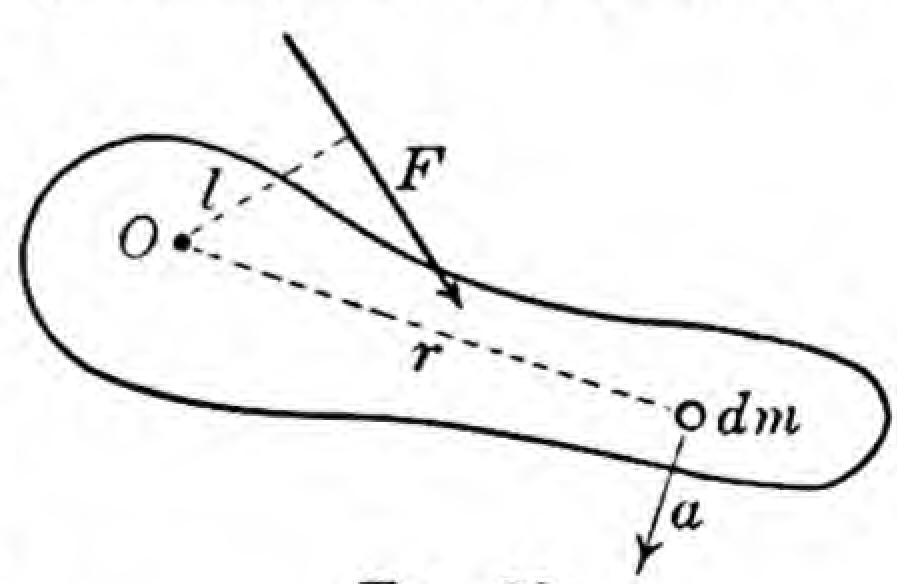


FIG. 39

Let a line perpendicular to the paper through the point O (Fig. 39) be the axis about which an irregular rigid body is free to rotate. Let a force F act on the body so that its lever arm with reference to O is l. The moment of the applied force is

$$L = Fl.$$

A mass dm, whose distance from O is r, is given a linear acceleration $a = \alpha r$. So the mass dm must be acted upon by a force (by

means of the connecting system) equal to $dm\ a = \alpha r\ dm$. The moment of the force about the axis through O is $\alpha r^2\ dm$. If we integrate this expression over the whole body, we get the sum of all the moments necessary to give each particle the required angular acceleration. But the total moment of force applied to the whole body which accomplishes that acceleration is L. So

$$L = \alpha \int r^2\ dm. \tag{70}$$

Comparing Eq. 70 with Eq. 69, we see that

$$I = \int r^2\ dm. \tag{71}$$

The moment of inertia of a given particle varies as the square of its distance from the axis of rotation. *It is always necessary to state where the axis is taken before the moment of inertia of a body may be stated.*

As an example, let us find the moment of inertia of a circular cylinder about an axis perpendicular to its face and through its center. Let the cylinder be of radius R, height h, and density ρ.

All elements of mass in a thin cylinder of radius r, thickness dr, and height h have the same radius r and may be taken together.

$$\begin{aligned} I &= \int_0^R dm\ r^2 = \int_0^R \rho\ dv\ r^2 = \int_0^R \rho(2\ \pi r\ dr\ h)r^2 \\ &= 2\ \pi\rho h \int_0^R r^3\ dr = \tfrac{1}{2}\ \pi\rho h R^4. \end{aligned}$$

The total mass of the disk,

$$M = \rho V = \rho\pi R^2 h.$$

So $I = \frac{1}{2} MR^2$. Note that the height does not appear explicitly in the final formula.

The following table gives the moment of inertia of several regular bodies. In each case the axis passes in a specified direction through the center of mass. M is the total mass of the body, l is the length, and R the radius.

After the moment of inertia of a body has been found, we may ask ourselves the question: If it were possible to have the whole mass of a body concentrated at a point or in a thin circular ring, at a certain distance from the axis, what would that distance

MOMENTS OF INERTIA
(Axis through center of mass)

Long thin rod	About a transverse axis	$\frac{1}{12} Ml^2$
Disk	Perpendicular to its plane	$\frac{1}{2} MR^2$
Thin disk	About a diameter	$\frac{1}{4} MR^2$
Cylinder	About a longitudinal axis	$\frac{1}{2} MR^2$
Cylinder	About a transverse axis	$M(l^2/12 + R^2/4)$
Sphere		$\frac{2}{5} MR^2$

have to be in order that the concentrated mass would have the same moment of inertia about that axis as has the actual rigid body? That distance is called the *radius of gyration* of the body about the axis considered and is designated by the letter k. Hence we have the defining equation,

$$I = Mk^2. \tag{72}$$

It is seen from this equation that k may be found by dividing the moment of inertia of the body by the total mass of the body and taking the square root of that quotient.

Since the above equations were derived by the application of Newton's second law, it will be necessary to replace M by W/g whenever Eq. 69 is used with engineering units. Thus in Eq. 69, when L is expressed in lb. ft. and α in rad./sec²., I must be computed by formulae such as in the above table with $M = W$ in lbs./32.2 ft./sec². and the dimensions of the body (which are always raised to the second power) must be expressed in ft.

In the metric system the torque L is expressed in cm. dynes, M in gms., the dimensions of the body in cm., and α in rad./sec².

Since the moment of inertia of a body is an arithmetic sum of terms for all the parts of the body, by breaking the summation into parts we see that the moment of inertia of a body about a given axis is the sum of the moments of inertia of each of its parts taken separately about the same axis. So if the moments of inertia of two separate bodies are known about a certain axis, then if these bodies are placed together so that the axes considered are coincident, the moment of inertia of the two bodies together is the sum of the moments of inertia of each separate body.

41. Moments of Inertia about Parallel Axes. — If the moment of inertia is known for any given axis passing through the center of mass, the moment of inertia about any axis parallel to the one given may be computed by a simple formula which will now be derived. Let yy' (Fig. 40) be any axis through the center of mass O. Let the moment of inertia about this given axis be I_0. The moment of inertia about a parallel axis distant h from yy' is desired. Let bb' be that axis. Through O draw the line Ox perpendicular to bb' and intersecting it at a. Take any element of mass dm distant r_1 from yy' and r_2 from bb'. Let the angle between r_1 and the xy plane be θ. Then by the Law of Cosines,

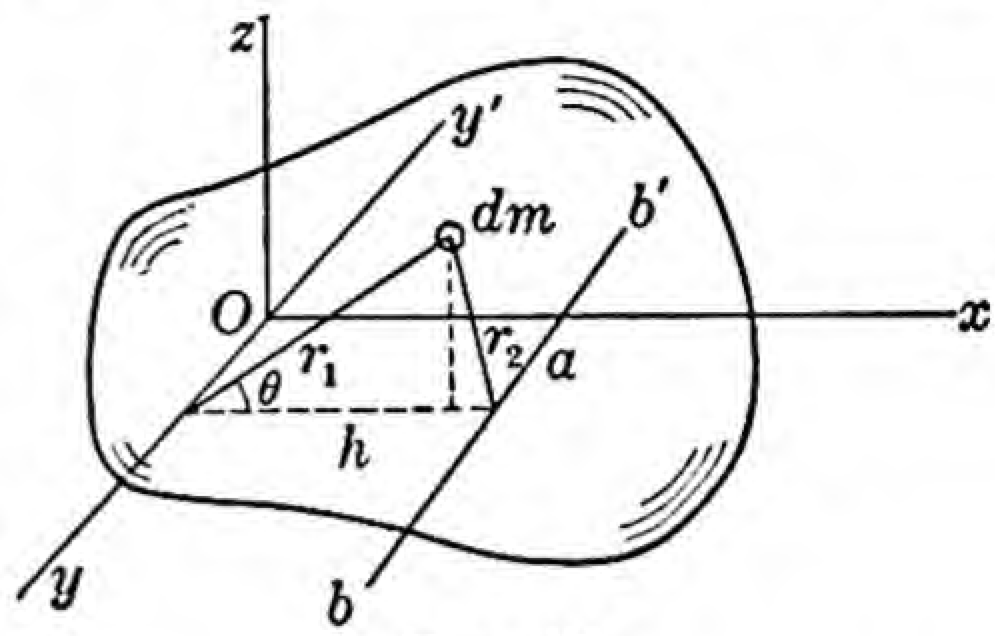


FIG. 40

$$r_2^2 = r_1^2 + h^2 - 2\,hr_1 \cos\theta.$$

Multiplying each term by dm and integrating so as to include the whole body, we have

$$\int r_2^2\,dm = \int r_1^2\,dm + h^2 \int dm - 2\,h \int r_1 \cos\theta \cdot dm.$$

It is apparent that $r_1 \cos\theta = x$, the x co-ordinate of dm, and

$$\int r_1 \cos\theta\,dm = \int x\,dm.$$

According to Eq. 63 this integral is equal to $\bar{x}M$. Since in this case we have chosen the axis through the center of mass, $\bar{x} = 0$ and therefore

$$\int r_1 \cos\theta\,dm = 0.$$

Hence
$$\int r_2^2\,dm = \int r_1^2\,dm + h^2 \int dm,$$

which may be written

$$I_a = I_0 + Mh^2. \tag{73}$$

It is to be carefully noted that I_0 is the moment of inertia about a given axis *passing through the center of mass of the body.*

42. Special Statement of Newton's Laws of Motion for Angular Motion. — In § 9, we showed that Newton's second law in its most general form is stated as

$$F\,dt = d(mv).$$

Similarly, for rotary motion we must write

$$L\,dt = d(I\omega). \tag{74}$$

The change of angular momentum of any body is directly proportional to the resultant moment of force acting upon it and to the time during which the moment acts and it takes place about the same axis as that of the moment of force. Eq. 69 is identical to Eq. 74 as long as the moment of inertia remains constant. Suppose that, without any force from the outside, the body in some way changes its shape during rotation. Both I and ω may thus change simultaneously. Let the first state be when the moment of inertia is I_0 and the angular velocity ω_0; the second state when the moment of inertia is I and the angular velocity is ω. From Eq. 74 we have,

$$L\,dt = d(I\omega).$$

Integrating, $$\int L\,dt = \int_{I_0\omega_0}^{I\omega} d(I\omega) = I\omega - I_0\omega_0. \tag{75}$$

From Eq. 75 we may state the second law as: The change in angular momentum of a body is equal to the impulse of the resultant moment of force and takes place in the same direction as the moment of force.

Suppose now that we consider the case of a rotating body on which the resultant torque is zero. Eq. 75 gives

$$I_0\omega_0 = I\omega. \tag{76}$$

We may then state Newton's *first law* of motion as applied to rotation: *If the resultant moment of force acting on a rigid body is zero, the angular momentum is conserved (does not change). Or, if a rigid body is at rest it remains at rest, or if in motion, it continues to rotate about the same axis with the same angular momentum.* While a body is rotating with a given angular velocity, if the body were allowed in some way to double its moment of inertia, then its angular velocity would decrease to half its former value and conversely.

The special statement of Newton's Third Law for rotary motion is left to the student.

43. Work, Power, and Energy in Angular Motion. — Let us consider the work done when a rigid body is rotated under the action of a torque. Let F be a force applied to the body at a point distant r from the axis of rotation O (Fig. 41). Let ϕ be the angle between F and ds. The work done by the force F in turning the body through a small angle $d\theta$ is

$$dW = F \cos \phi \, ds = F \cos \phi \, r \, d\theta = L \, d\theta.$$

For a finite displacement θ, the total work is

$$W = \int_0^\theta L \, d\theta, \tag{77}$$

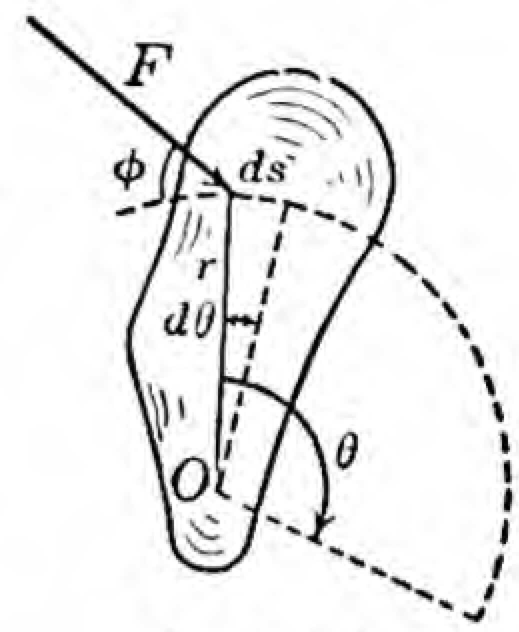


FIG. 41

where L is the applied torque. If there are several forces acting on the body, then L represents the resultant torque. This expression is integrable only after L and θ are expressed in a common variable. If the applied torque is constant, then we obtain $W = L\theta$, where θ is the total angle through which the body is turned under the action of the torque L. It is to be seen from the steps in the proof that if F is in dynes, r in cm., and θ in radians, W will be in ergs. If F is in lbs., r in ft., and θ in radians, W will be in ft. lbs.

The power or the rate at which work is done by the torque is given by the expression

$$\text{Power} = \frac{dW}{dt} = \frac{L \, d\theta}{dt} = L\omega, \tag{78}$$

where L represents either the torque due to a single force or the resultant torque of a set of forces. If either L or ω or both are variable, then Eq. 78 gives only the instantaneous power.

Let a resultant moment of force, either constant or variable, act upon a body whose moment of inertia I is constant and whose angular velocity is changed from ω_0 to ω. By steps similar to those by which Eq. 37 was developed, we find that the work done on the body is

$$W = \tfrac{1}{2} I(\omega^2 - \omega_0^2). \tag{79}$$

The derivation is left for the student. At what step is it required that I must be constant?

In Eqs. 78 and 79 the student should show that if ω is expressed in rad./sec., L in lb. ft., and I in slug ft^2., then P is in ft. lbs./sec. and W in ft. lb.; if L is in cm. dynes and I in gm. cm^2., then P is in ergs/sec. and W in ergs.

44. Motion of a Rolling Body. — In the case of a rolling sphere or circular cylinder, the motion may be considered (*a*) as a composition of a pure rotation with a certain angular velocity ω_0 about an axis through the center of the body, and a pure translation of the center of the body, or (*b*) as a pure rotation about an instantaneous axis passing through the surface of contact, with an angular velocity which we shall prove to be equal to ω_0.

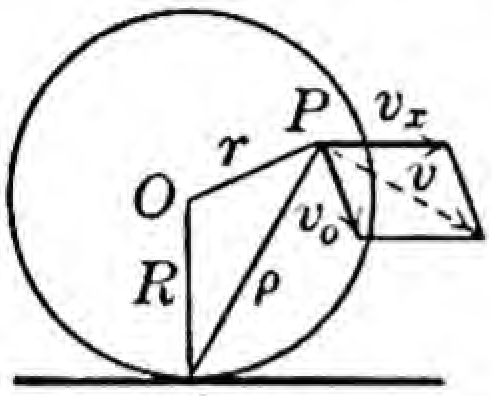


FIG. 42

In Fig. 42 the point P, distant r from the center O, may be considered as having a velocity component v_x which is the same as the velocity of the center and a component $v_0 = \omega_0 r$ due to the rotation of the body with angular velocity ω_0 about O. The velocity v, the resultant of these components, will now be proved to be perpendicular to AP. Also ω_A, the angular velocity of point P about A, will be shown to be equal to ω_0.

The point A is temporarily at rest. Its velocity forward, v_x, equal to that of the center O, must be equal to its velocity backward, $\omega_0 R$, due to its rotation about O. So $v_x = \omega_0 R$. Also we know that $v_0 = \omega_0 r$.

Hence
$$\frac{v_x}{v_0} = \frac{R}{r}.$$

The student should show that it follows therefore that the triangle AOP is similar to either vector triangle and that therefore v is perpendicular to ρ. We may therefore consider that the point P is rotating about the point of contact. From the similar triangles we obtain,

$$\frac{v_0}{r} = \frac{v}{\rho}$$

or
$$\omega_0 = \omega_A.$$

Hence any point in the rolling body may be considered to be rotating about an axis through the instantaneous point of contact with an angular velocity equal to that about the axis through the center.

Obviously if the body is accelerated, then $\alpha_0 = \alpha_A$.

45. Energy Relations for a Rigid Body Rolling on an Incline. — Let a sphere or cylinder of radius R roll from rest down a plane of length s and height h (Fig. 43). When the body reaches the bottom, it has a linear velocity v and an angular velocity ω_0 about O. It has lost potential energy of amount mgh which must appear as kinetic energy.

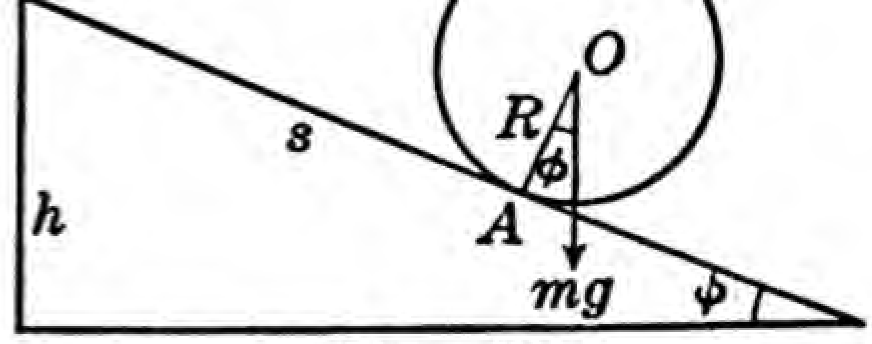


FIG. 43

Referring to the point O, we say that at the bottom of the incline the body has kinetic energy $\frac{1}{2}mv^2$ due to its linear velocity and $\frac{1}{2}I_0\omega_0^2$ due to its rotation about O. I_0 is the moment of inertia about O. $I_A = I_0 + mR^2$. Then

$$mgh = \frac{1}{2}mv^2 + \frac{1}{2}I_0\omega_0^2 = \frac{1}{2}I_A\omega_0^2. \tag{80}$$

The last term of this equation shows again that we may consider the point A to be instantaneously at rest and the body to have only pure rotation about it.

Likewise we may consider that the kinetic energy is produced by the gravitational force mg which causes rotation about the point A. The torque is $mgR \sin \phi$. As the body rolls down the plane the torque acts through an angle $\theta = s/R\cdot$ Remembering that $\sin \phi = h/s$, the student should show that the work done by the torque reduces to the expression mgh.

PROBLEMS

1. Find the moment of inertia of a sphere of radius R about any axis passing through its center. Take the y axis as the axis of rotation and consider the sphere made up of disks of radius $\sqrt{R^2 - y^2}$ and thickness dy. Get the value of I for this disk of infinitesimal thickness by use of the formula in § 40 and integrate this expression from $-R$ and $+R$.

2. A circular disk of radius R_2 has a smaller disk of radius R_1 cut out of its center, leaving a ring of rectangular cross section. Using the formula for the moment of inertia of a disk, show that if M is the mass of the ring, the moment of inertia of the ring about an axis normal to its plane and through its center is

$$\frac{1}{2}M(R_1^2 + R_2^2).$$

3. Find the radius of gyration of each of the bodies listed in the table in § 40.

4. Find the moment of inertia of

(*a*) A long thin rod about an axis perpendicular to its length, (1) passing through its end and (2) passing through its center.

(*b*) A thin circular disk about its diameter.

(*c*) A thin rectangular plate about an axis through its center and bisecting one set of sides.

5. Using the parallel axis theorem, correlate the formulae already derived for the moments of inertia of a rod about axes through the center and one end.

6. In Problem 4 (*c*), if the plate is twice as long as it is wide, what is the relative increase in its moment of inertia when the axis which bisected the short sides is changed so as to bisect the long sides?

7. A horizontal axle weighing 300 lbs. and 2 inches in diameter has mounted on it a solid disk flywheel weighing 200 lbs. and 1 foot in diameter.

(*a*) What torque, in addition to that necessary to balance out friction, is necessary to give the flywheel an angular velocity of 20 radians per second in 2 seconds?

(*b*) When revolving at that rate the torque is removed and friction brings the system to rest after it has turned through 600 radians. What was the total torque required in part (*a*)?

(*c*) What power, including friction, is required to maintain a constant angular speed of 30 rad./sec.?

(*d*) In part (*a*) find the applied power at the end of the first second.

(*e*) A cord is wound around the rim of the flywheel and a 10 lb. weight hung onto the free end of the cord. Starting from rest, how long will it take the weight to drop 10 ft.? When it has fallen that distance, what is its linear velocity, the angular velocity of the axle, the kinetic energy of the whole system, and the tension in the cord.

8. A sphere, a cylinder, and a cube, each weighing 1 kg., start from rest and roll (or slide) down a frictionless 30° incline. The radii of the sphere and the cylinder are the same. The incline is 10 meters long. Which body reaches the bottom first, what is the time of descent of each, and for each body how is the energy distributed between that of translation and rotation? What would be the effect on these answers if the weights were all doubled or if they were not the same for the three bodies?

9. The bar and drum shown in Fig. 44 have a moment of inertia of 10 slug ft². and the system is given an angular velocity of 360 r.p.m. After the power is removed what distance will the suspended mass rise? What is the time of ascent? What is the tension in the cord when the mass is rising and what is it later when the mass is falling?

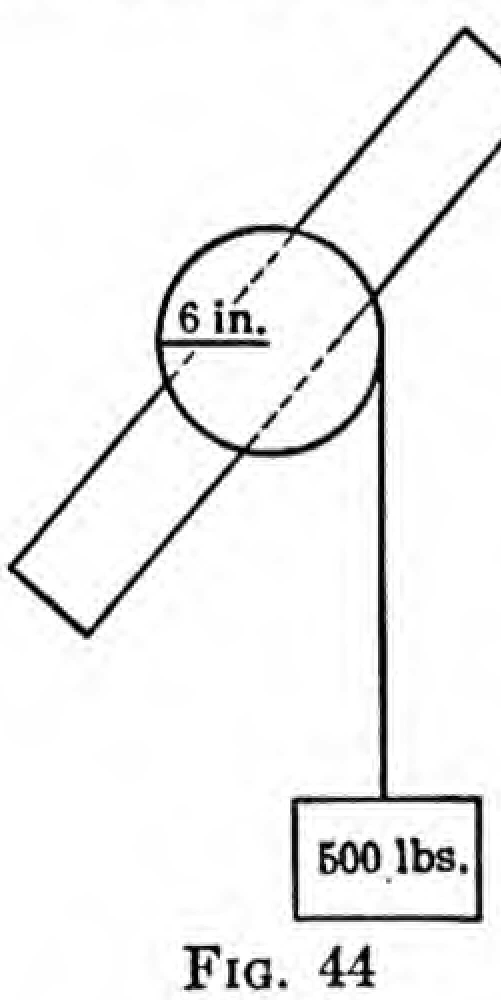


FIG. 44

10. A steel shaft 40 feet long and 4 inches in diameter transmitting 100 horsepower makes 14 r.p.s. What is the torque? If the slide modulus of steel is 12×10^6 pounds per square inch, calculate the angle through which one end of the shaft is twisted relative to the other end. (See Eq. 78.)

11. (*a*) If the steel shaft in Problem 10 were to transmit 150 H.P. with the same torque, at what speed must the shaft turn?

(*b*) If the shaft were to transmit 150 H.P. at the same speed, what would be the torque and through what angle would one end of the shaft be twisted relative to the other end?

12. The moment of inertia of a body is 7000 gm. cm². If a torque of 35,000 dyne cm. is applied to the body, find the angular velocity at the end of 8 seconds if the initial velocity is 12 radians per second. What will be the change in kinetic energy?

13. A horizontal flat circular disk of 3 ft. diameter weighs 100 lbs. Two spheres of 1 ft. diameter, weighing 50 lbs. each, are placed on a diameter of the disk on opposite sides of its center. The centers of the spheres are each one foot from the center of the disk. What is the moment of inertia of the system about a vertical axis through the center of the disk?

14. Prove that the potential energy stored in a twisted shaft is equal to $\frac{1}{2} L_0\theta^2$, where L_0 is the torque required to twist the shaft through one radian and θ is the angle through which the end of the shaft is turned.

Gyroscopic Motion

46. The Vector Addition of Angular Velocities. Precession. — Angular velocity is a physical quantity having both magnitude and direction and therefore may be represented by a vector. The vector direction is always taken parallel to the axis of rotation of the body being considered, the sense of the direction being taken as that in which a right-handed screw would advance if given the same motion. Angular velocities are additive according to the laws of vector addition. Suppose that a body is rotating about an axis parallel to the x axis (Fig. 45), with an angular velocity ω_1 represented by the vector OA. We wish to find the result of giving the body an angular velocity about another axis. Suppose that we give the body an additional angular velocity $d\omega$ which is parallel to the y axis,—or, as viewed, raise the right-hand side of the body and lower the left side. The sum of the two velocities, $\omega_1 + d\omega$, is ω_2, which is represented by line OB. The result is that the direction of the axis of rotation of the body is changed. The rotation has taken place about the z axis. Since $d\omega$ is small, then $\omega_2 = \omega_1$. Now let another small vector $d\omega$ perpendicular to ω_2 be added to ω_2, etc. The result is a continuous rotation of the axis of rotation of the body about the z axis, leaving the magnitude of the velocity ω_1 unchanged. Thus when a body is rotating about an axis and a torque is applied tending to produce rotation about some other axis perpendicular to that axis, the original axis of rotation is turned about a third axis perpendicular to the other two axes. This is the essence of gyroscopic motion. We shall apply it to one of the common forms of gyroscope, a wheel of considerable moment of inertia supported in a universal mounting (Fig. 46), so that it is free to rotate about any axis through its center. Let the wheel be rotating about its axis MN with the top moving toward the observer

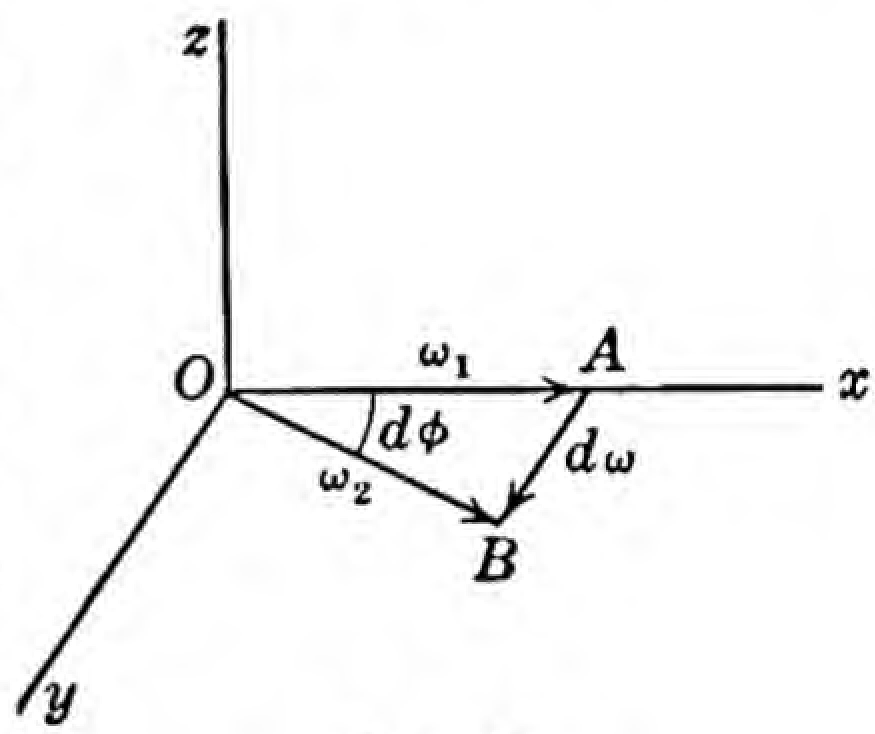


Fig. 45

Fig. 46

as represented by the vector OA (Fig. 45). Let a weight W be applied to point M. This produces a constant torque and according to Newton's second law will increase the angular velocity along the axis PQ by an amount $d\omega$ during every equal interval of time dt. The result will be that the axis MN rotates at a constant rate about the line SR, point N moving from its position shown toward the observer. This motion is called precession. Since the weight moves with the axis, its torque is always producing a change in angular velocity at right angles to the existing angular velocity about MN, so the original velocity remains unchanged in amount as explained above. It is rather astonishing at first that the weight W does not cause the point M to sink and N to rise. However, such a motion would violate the second law of motion because the torque produced by W must cause an increase in angular momentum only in the same direction in which it acts. It acts along PQ, directed toward Q. Now the original angular momentum lies along MN, so both the original and the change in the angular momentum lie in the same plane. The addition of these quantities could not possibly give a vector lying out of the plane as would be necessary if the point M were lowered and N raised by the application of the weight.

47. Quantitative Relations. — In Fig. 45 we see that if the torque is applied to a body rotating for an infinitesimal time dt, the increase in angular velocity $d\omega$ is infinitesimal and $\omega_1 = \omega_2 = \omega$, the original angular velocity. Let $d\phi$ be the angle through which the original axis of rotation changes. Then $d\omega = \omega\, d\phi$ and

$$\alpha = \frac{d\omega}{dt} = \omega \frac{d\phi}{dt}.$$

The expression $d\phi/dt$ gives the angular velocity with which the original axis precesses and it will be called ω'. Let the lever arm of the applied force be R. Then by Newton's second law,

$$L = WR = MgR = I\alpha = I\omega\omega'. \tag{81}$$

According to this expression we see that if W is doubled, the velocity of precession is doubled; if the moment of inertia of the wheel is doubled and the wheel has the same angular velocity ω as before, then W must be doubled or R doubled in order to have the same precessional velocity; for a given W and I, if the wheel is spinning with twice as large an angular velocity, the precessional velocity will be only half as large; etc.

Harmonic Motion

48. Definition of Simple Harmonic Motion. — Simple harmonic motion is motion in which the magnitude of the acceleration is directly proportional to the displacement from the position of equilibrium but the direction of the acceleration is opposite that of the displacement. From Newton's second law and Hooke's law, we see that, when an elastic body is distorted and released, this type of motion will result. Common examples of this type of motion are the vibration of stretched strings or drum heads, twisted wires, bent bars, etc. By cams and other devices bodies are often forced to move with simple harmonic motion.

For the case of motion of a particle in a straight line the defining equation of S.H.M. is

$$a = -\omega^2 s, \text{ or } \frac{d^2s}{dt^2} = -\omega^2 s. \tag{82}$$

The reason for choosing ω^2 as the proportionality factor will be obvious when the final solution is obtained. This is a very common type of differential equation of second order. The most general solution must be an expression for s in terms of t and containing two arbitrary constants. If any such expression is a solution, then when its values are substituted in Eq. 82 an identity must result, *i.e.* the solution must satisfy the original differential equation.

49. Solution of the Equation of S.H.M. — Multiplying both sides of Eq. 82 by $2\,ds/dt$,

$$2\frac{ds}{dt}\frac{d^2s}{dt^2} = \frac{d}{dt}\left(\frac{ds}{dt}\right)^2 = \frac{d(v)^2}{dt} = -2\,\omega^2 s\,\frac{ds}{dt}.$$

Integrating this expression,

$$v^2 = \left(\frac{ds}{dt}\right)^2 = -\omega^2 s^2 + c.$$

The velocity of the particle reverses direction at the maximum displacement, s_m, and, therefore, goes through the value of zero at that point. So it is convenient to select as the initial conditions, $v = 0$, $s = s_m$. Then we find that $c = \omega^2 s_m{}^2$ and the above equation becomes

$$v = \frac{ds}{dt} = \pm\,\omega\sqrt{s_m{}^2 - s^2}, \tag{83}$$

or

$$\pm\frac{ds}{\sqrt{s_m{}^2 - s^2}} = \omega\,dt.$$

Integrating this expression, using first the positive sign and then the negative sign, we obtain, respectively,

$$\sin^{-1}\frac{s}{s_m} = \omega t + e, \tag{84}$$

or

$$\cos^{-1}\frac{s}{s_m} = \omega t + e'.$$

Hence

$$s = s_m \sin(\omega t + e), \tag{85s}$$

or

$$s = s_m \cos(\omega t + e').$$

The values of e and e' are determined by what the displacement is when we start counting time, *i.e.* what value s has when $t = 0$.

Substituting the values of s from Eq. 84 into Eq. 83, the student should obtain the following expressions for the velocity and acceleration of the particle.

$$v = \omega s_m \cos(\omega t + e), \tag{85v}$$

or

$$v = -\,\omega s_m \sin(\omega t + e'),$$

and

$$a = -\,\omega^2 s = -\,\omega^2 s_m \sin(\omega t + e), \tag{85a}$$

or

$$a = -\,\omega^2 s_m \cos(\omega t + e').$$

50. The Projection of Uniform Circular Motion. — Consider a particle A moving around a circle with uniform speed. Let this motion be projected onto any straight line, such as MN (Fig. 47). Let displacements along MN be measured from the point O, the projection of the center O' of the circle. The motion consists of displacements between the values $\pm\, s_m$. The radius of the circle is s_m. Let us measure the angular position of the point A by the angle between $O'A$ and $O'O$. Let ω be the angular velocity of the line OA and let time be counted from the instant when the particle passes through the point B. Call angle $OO'B$, e. For the displacement, s, from O we get

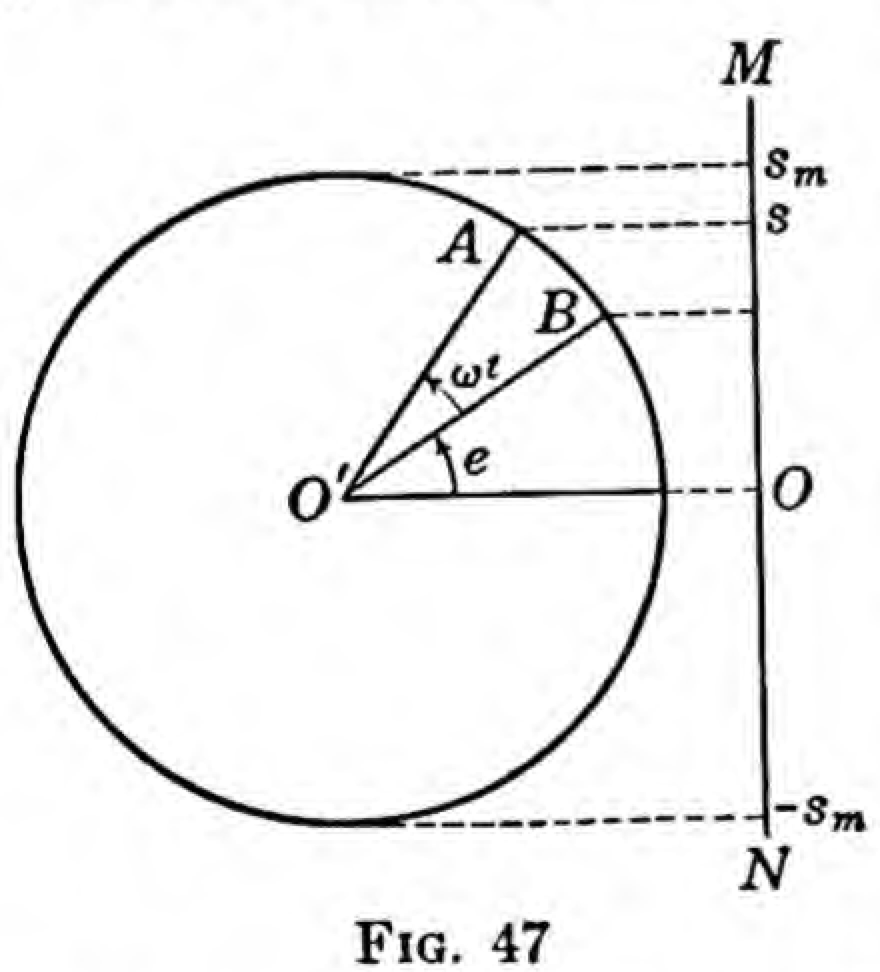


FIG. 47

$$s = s_m \sin(\omega t + e).$$

If we had chosen a line perpendicular to MN onto which to project the motion and still chosen $O'O$ for a measure of the angles, we would have got

$$s = s_m \cos(\omega t + e').$$

Therefore it follows from our definition in Eq. 82 that linear S.H.M. is the projection of uniform circular motion on a straight line. It is now seen why in § 49 the solution turned out to be a trigonometric expression involving angles although we were considering motion on a straight line.

51. Physical Meaning of the Solution. — The solution expresses the displacement as a sine or a cosine function of an angle which varies directly with the time. The sine or cosine of such an angle goes through a certain definite cycle of values after which it begins to repeat itself. The maximum value of a sine or cosine is unity; hence s_m is the maximum value of the displacement or is the *amplitude* of the motion. $(\omega t + e)$ or $(\omega t + e')$ is called the phase angle of the motion. It is the angle whose sine or cosine is the ratio of the displacement at any instant to the maximum displacement. e or e' is the phase angle at zero time. The motion begins to repeat itself when $\omega t = 2\pi$ or when $t = 2\pi/\omega$. This particular value of t is called the *period* T of the motion. Therefore, $\omega^2 = 4\pi^2/T^2$ and $\omega = 2\pi/T$. Hence ω is the angular velocity of the point A on the reference circle. The physical significance of each symbol in the Eq. 85s is thus known.

As pointed out in the previous paragraph, it is a matter of choice as to whether the solution is used in the sine or cosine form. In order that both equations represent the motion of the same particle, obviously e' would have to be different from e by 90°. It is customary to use the sine form of the solution whenever possible.

52. The Converse Proof. — We shall now show that if we define S.H.M. as the projection on a straight line of uniform circular motion, it follows that the motion is such that the acceleration is directly proportional to the displacement and oppositely directed.

As a matter of convenience, throughout the remainder of the discussion, we shall count time from the instant when the particle passes through O (Fig. 48), the center of its path. Then $e = 0$. When the particle is at P, the point of the reference circle is at A. Then angle $OO'A$ is

$$\omega t = \frac{2\pi}{T}t.$$

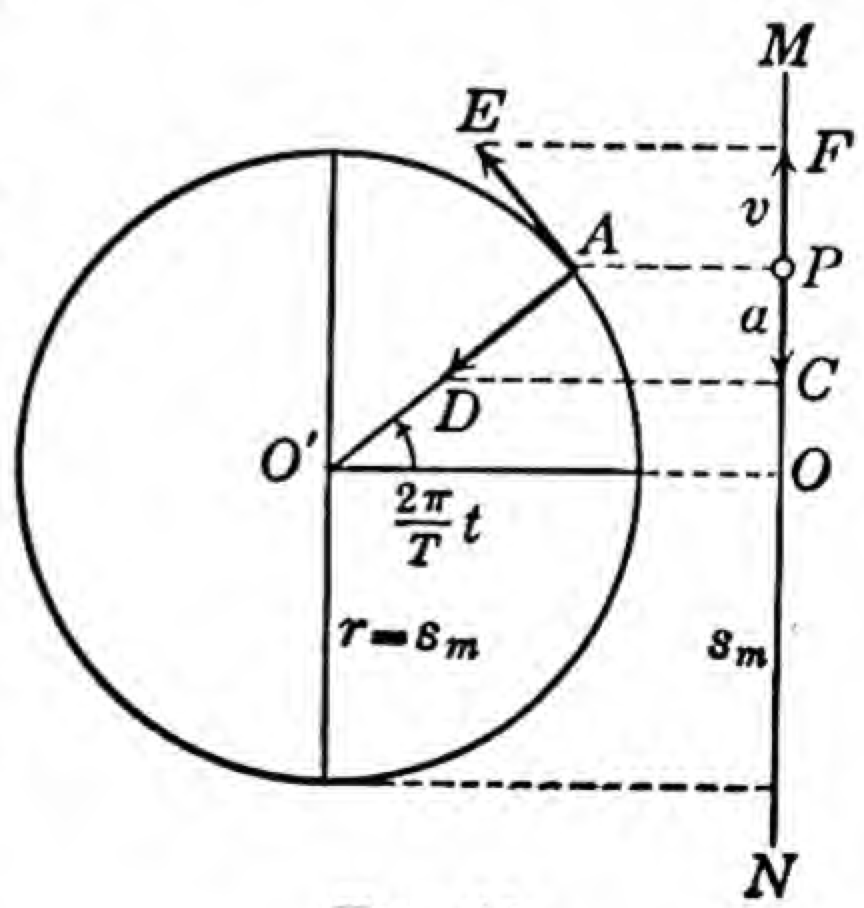


FIG. 48

The linear velocity of A is

$$AE = \omega r = \frac{2\pi}{T} s_m.$$

The acceleration of A is

$$AD = \frac{v^2}{r} = \omega^2 r = \frac{4\pi^2}{T^2} s_m.$$

The velocity and acceleration of the point P are, respectively, the projection of AE and AD upon MN, namely PC and PF. Hence we have for OP, PF, and PC the following equations:

$$s = s_m \sin \frac{2\pi}{T} t, \tag{86s}$$

$$v = \frac{2\pi}{T} s_m \cos \frac{2\pi}{T} t, \tag{86v}$$

and

$$a = -\frac{4\pi^2}{T^2} s_m \sin \frac{2\pi}{T} t. \tag{86a}$$

The negative sign must be placed before the last term because we see from the figure that as long as s is directed upward the acceleration is directed downward, and conversely.

By inspection of the equations for s and a we see that they differ by only the factor

$$-\frac{4\pi^2}{T^2} = -\omega^2.$$

Hence we have, as was to be proved,

$$a = -\omega^2 s.$$

The equations for the velocity and acceleration may be easily obtained by differentiating the expression for the displacement. Thus,

$$s = s_m \sin \omega t = s_m \sin \frac{2\pi}{T} t, \tag{87s}$$

$$v = \frac{ds}{dt} = \frac{2\pi}{T} s_m \cos \frac{2\pi}{T} t = v_m \cos \frac{2\pi}{T} t, \tag{87v}$$

and

$$a = \frac{d^2 s}{dt^2} = -\frac{4\pi^2}{T^2} s_m \sin \frac{2\pi}{T} t = -\frac{4\pi^2}{T^2} s. \tag{87a}$$

53. The Period of Simple Harmonic Motion. — From the last of these equations we obtain

$$T = 2\pi\sqrt{-\frac{s}{a}}. \tag{88}$$

Since s and a are vectors and always of opposite sign, the expression under the radical is essentially positive and for this reason the sign is sometimes omitted. The above equation may also be derived by using two previous equations: $a = - \omega^2 s$ and $\omega = 2\pi/T$.

Considering the mass of the vibrating particle and the force acting on it, $F = ma$, Eq. 88 becomes

$$T = 2\pi \sqrt{\frac{m}{\frac{F}{s}}}. \tag{89}$$

This latter formula is very useful. It enables one to find the period of vibration when the force for a given displacement is known or to find the force necessary to produce unit displacement by observing the period of the motion. If F is to be in gravitational units, m must be replaced by W/g, and s must be in ft. If m is in gms., F must be in dynes, and s in cms. T must always be in seconds.

From Eqs. 87s and 87v, we see that s and v are always 90° out of phase, while s and a are always 180° out of phase. These relations are shown analytically in Fig. 49.

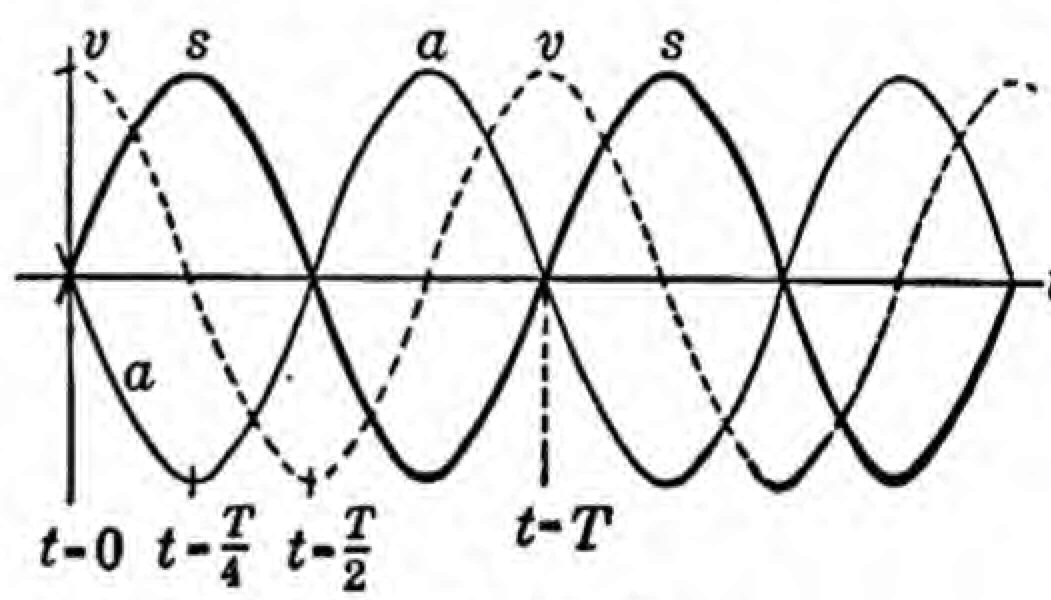


FIG. 49

When s is at its maximum, a is at its maximum in the opposite direction, and v is zero.

When s is zero, a is zero, and v is at its maximum.

When s is positive, a is negative and conversely.

54. Angular Simple Harmonic Motion. — So far we have considered only linear Simple Harmonic Motion. Now let us discuss the case of angular Simple Harmonic Motion. One example of this is the so-called torsion pendulum. A heavy disk is supported by an elastic wire (Fig. 50). The disk is rotated through an angle θ about an axis through the wire and then released. At every instant of time, by Hooke's law, the return torque is directly proportional to the angular displacement θ and is oppositely directed to it. By Eq. 69, the angular acceleration is directly proportional to the return torque, so we have

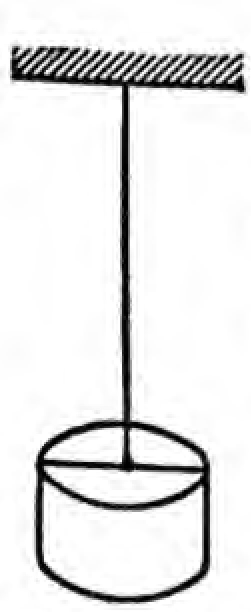


FIG. 50

$$\alpha = - k^2\theta, \text{ or } \frac{d^2\theta}{dt^2} = - k^2\theta. \tag{90}$$

This equation is identical in mathematical form to Eq. 82, so the solutions of the two equations will be identical in form. We need only to place L for F, I for m, α for a, θ for s, and ω for v. We then have the following equations:

$$\alpha = -k^2\theta,$$

$$\theta = \theta_m \sin \frac{2\pi}{T} t, \tag{91θ}$$

$$\omega = \frac{2\pi}{T} \theta_m \cos \frac{2\pi}{T} t, \tag{91ω}$$

$$\alpha = -\frac{4\pi^2}{T^2} \theta_m \sin \frac{2\pi}{T} t = -\frac{4\pi^2}{T^2} \theta, \tag{91α}$$

$$T = 2\pi\sqrt{\frac{\theta}{\alpha}} = 2\pi\sqrt{\frac{I}{\frac{L}{\theta}}} = 2\pi\sqrt{\frac{I}{L_0}}. \tag{92}$$

The amount of torque required to twist a wire through one radian, L/θ or L_0, is called the *torsion constant* for the wire.

Eq. 92 is useful in determining experimentally the moment of inertia of an irregular body. Let the disk shown in Fig. 50 be regular and accurately made. Its moment of inertia, say I_1, may be computed from its weight and its dimensions. Let the observed period be T_1. Then the irregular mass is supported on top (or bottom) of the disk with the axis about which the moment of inertia is desired coinciding with that of the torsion wire. Let I_2 be the unknown moment of inertia. We have the equations

$$T_1 = 2\pi\sqrt{\frac{I_1}{L_0}},$$

and

$$T_2 = 2\pi\sqrt{\frac{I_1 + I_2}{L_0}}.$$

Then

$$\frac{T_1^2}{T_2^2} = \frac{I_1}{I_1 + I_2}.$$

From this latter equation the value I_2 may be computed.

Often we wish to know the value of a small torque which has caused a fine elastic wire to twist a certain measured amount. We may suspend at the end of the wire a regular disk which the wire will support without being strained beyond its yield point. Its period as a torsion pendulum is measured. The value of I for the disk is computed. Then L_0 is computed from Eq. 92. Since L_0 is the torque which will produce the deflection of one radian, the desired torque which caused the given twist is determinable.

PROBLEMS

1. In the case of a stretched or compressed spring the return force is known to be proportional to the distortion and in the opposite direction, that is, $F = -cs$, where c is known as the spring constant. Define c in physical terms. Show that if a mass m is attached to the spring, it will vibrate with a period expressed by the equation

$$T = 2\pi\sqrt{m/c}. \tag{93}$$

2. Find the expression for the period of a simple pendulum. What is the effect of the mass on the period in this case? This problem will have to be solved approximately, — for the case where the swing is through such small angles that $\sin\theta$ is approximately equal to θ.

3. A mass of 100 gms. hanging on a spring is displaced 4 cm. and the return force is then 5000 dynes. If the mass is displaced 20 cm. and then released, find the period, the maximum velocity, and the maximum acceleration.

4. The period of a body undergoing Simple Harmonic Motion is observed. What effect will doubling the amplitude of the motion have upon the period? Can you make the result seem plausible?

5. A mass of 10 lbs. is hung on the end of a spiral spring. When set vibrating, it is found to have a period of $\frac{1}{2}$ sec. Find the "constant" of the spring and the force required to stretch the spring 3 in.

6. A cart weighing 1 kg. has frictionless wheels and is held on a 30° inclined plane by a spiral spring (Fig. 51). The spiral spring, when hung freely, is stretched 2 cm. by a 100 gm. weight. Find the period with which the cart on the incline will vibrate when displaced.

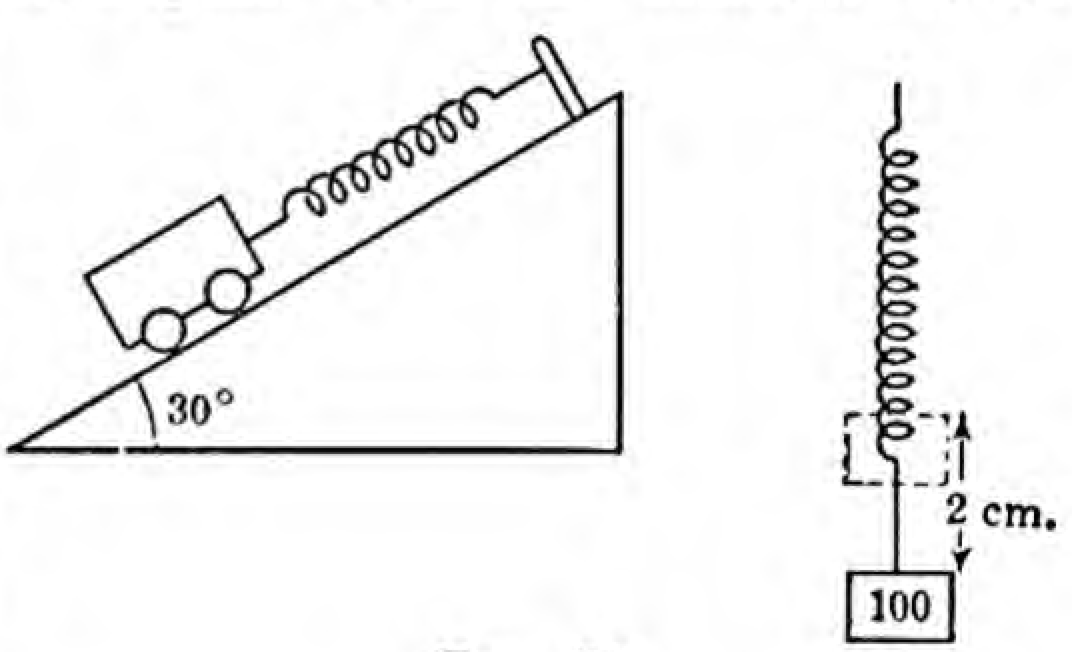


FIG. 51

7. A stick of cross section 4 cm². is loaded with lead on one end so as to float upright in water. The mass of the stick and the lead is 250 grams. When lowered or raised in the water from its equilibrium position, it will oscillate up and down. Will the motion be Simple Harmonic? Neglecting friction and the inertia of the water set into motion, find the period.

8. A body vibrates with Simple Harmonic Motion in a straight line. Find the average displacement during a half period (starting from the rest position). May the average force be found directly from this result?

9. Given a long spiral spring, a standard mass, a meter stick, and a stop watch, describe how an experiment might be performed to determine the gravitational acceleration. Why must a correction always be applied on account of the finite mass of the spring?

10. In the experiment described in the latter part of § 54, a disk 8 cm. in diameter, 1 millimeter thick, and weighing 15 gms. was supported by a wire. The period was 1.65 seconds. An irregular disk was added and the period

was then 2.06 seconds. What was the moment of inertia of the irregular disk and what was the torsion constant of the wire?

11. Two spiral springs are suspended from a support. On one of them is twice the weight of that on the other. The periods of each are the same. What may be said of the relative strengths (F/s) of the springs?

12. Any rigid body suspended so as to swing freely about an axis is called a compound physical pendulum. Let the moment of inertia about the axis of support be I and the distance from the axis to the center of gravity of the body be h. Prove that the period of the pendulum for small angles is

$$T = 2\pi\sqrt{\frac{I}{mgh}}. \tag{94}$$

13. A mass of 156 gms. is suspended by a spring such that an additional 40 gms. causes a further displacement of 2 cm. The total mass (196 gms.) is pulled down 2 cm. and released. Compute the period. Find the direction and magnitude of the displacement, velocity, and acceleration, one-half second after the mass passes downward through its equilibrium position.

14. A tooth in the blade of a reaper describes Simple Harmonic Motion of 1.5 in. amplitude in a period of 0.2 sec. Find its velocity when at a point 1.0 in. from the center of its path.

15. A grindstone is rotating with a uniform velocity of 60 r.p.m. Light from the afternoon sun casts a shadow of the handle on the ground. The sun is in the plane of the wheel but is 30° from the vertical. The handle is 1 ft. from the axis of rotation. Find the velocity and acceleration of the shadow when the handle is at its lowest point.

16. Mercury in a U-tube is disturbed. Show that, neglecting friction, the liquid executes a Simple Harmonic Motion of period $T = 2\pi\sqrt{l/2g}$, where l is the length of the liquid from surface to surface around the bend.

17. What would Eqs. 87 become for the following cases: $t = 0$ when the displacement is a maximum positive value; maximum negative value. $t = 0$ when the displacement is half of the maximum positive value; one half the maximum negative value.

CHAPTER II

WAVE MOTION AND SOUND

55. Simple Harmonic Motion and Wave Motion. — In elastic bodies, whenever a small particle is displaced, either by stretch, compression, or shear, there is a return force directly proportional to the displacement and when the displaced particle is released the resulting motion is simple harmonic. Not alone does the displaced particle vibrate, — it cannot vibrate without displacing the neighboring particles and thus causing them to vibrate. How quickly the motion of a particle is transferred to the neighboring particles depends upon the modulus of elasticity and the density of the material. If the modulus is large, then the intermolecular forces, for a given displacement, will be relatively large and the adjacent particles will accelerate rapidly. Also if the density of the body is small, then the forces will be able to accelerate the particles of small mass more rapidly. Thus the disturbance will travel rapidly in a medium of large modulus of elasticity and low density and conversely. If a given particle is kept in constant vibration by means of an external stimulus, then a continuous set of waves radiates from it as a source. Each particle in the body vibrates with simple harmonic motion, but due to the finite time taken for impulses to travel through the body, the particles more and more distant from the source are farther and farther behind in phase.

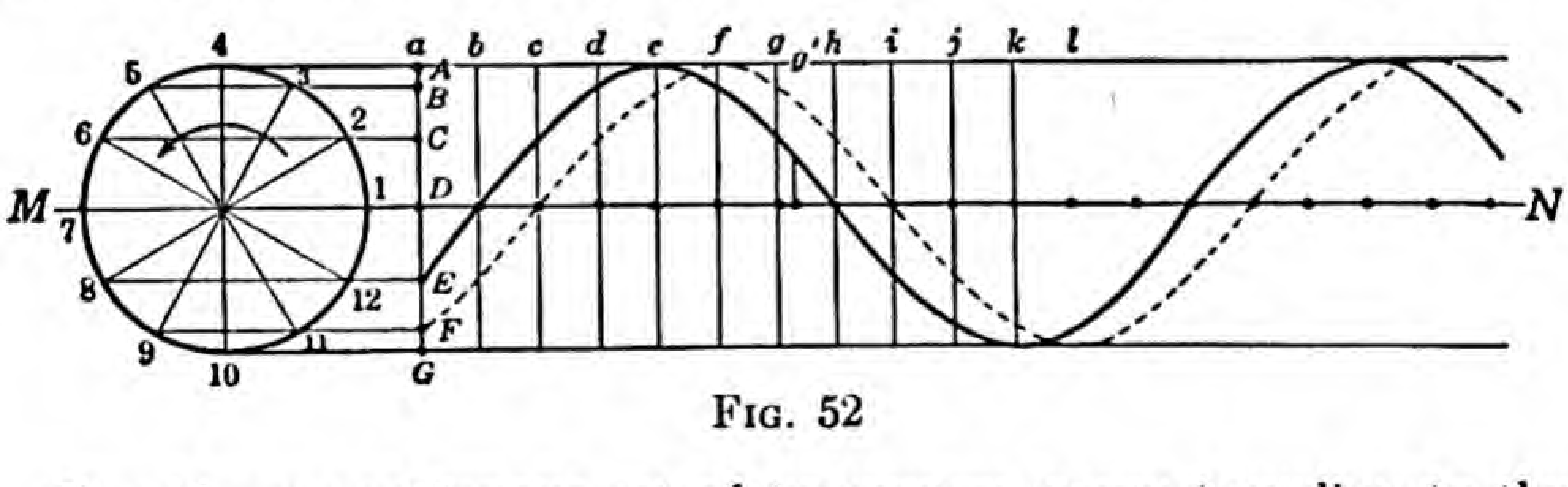


FIG. 52

Consider a continuous set of transverse waves traveling to the right along a cord MN (Fig. 52). It will be shown later that if the amplitude of the motion is not too large, each particle moves with simple harmonic motion vertically up and down at right angles to the direction of advance of the wave. For convenience, only the

particles *a*, *b*, *c*, etc., spaced one twelfth of a wave-length apart, will be followed. The position of each particle will be followed by positions 1 to 12 on the circle of reference placed at the left of Fig. 52. Let one configuration of the cord be represented by the heavy line and a later position by the dotted line. For the heavy line, the particle *a* is displaced downward by an amount *DE*, corresponding to the position 8 on the circle. The particle *b* is not as far advanced in phase as *a*. Its displacement is zero, corresponding to position 7. Likewise particle *c* is displaced upward an amount *DC* corresponding to position 6. Particle *d* is located by point 5, etc. One twelfth of a period later the displacement of any given particle will be given by a point on the reference circle advanced 30° from its former position; the displacement of particle *a* will be given by point 9, *b* by 8, *c* by 7, etc., the dotted line in Fig. 52 giving the positions of all particles.

Consider next a compressional wave traveling through an elastic medium. Each particle executes simple harmonic motion in a line

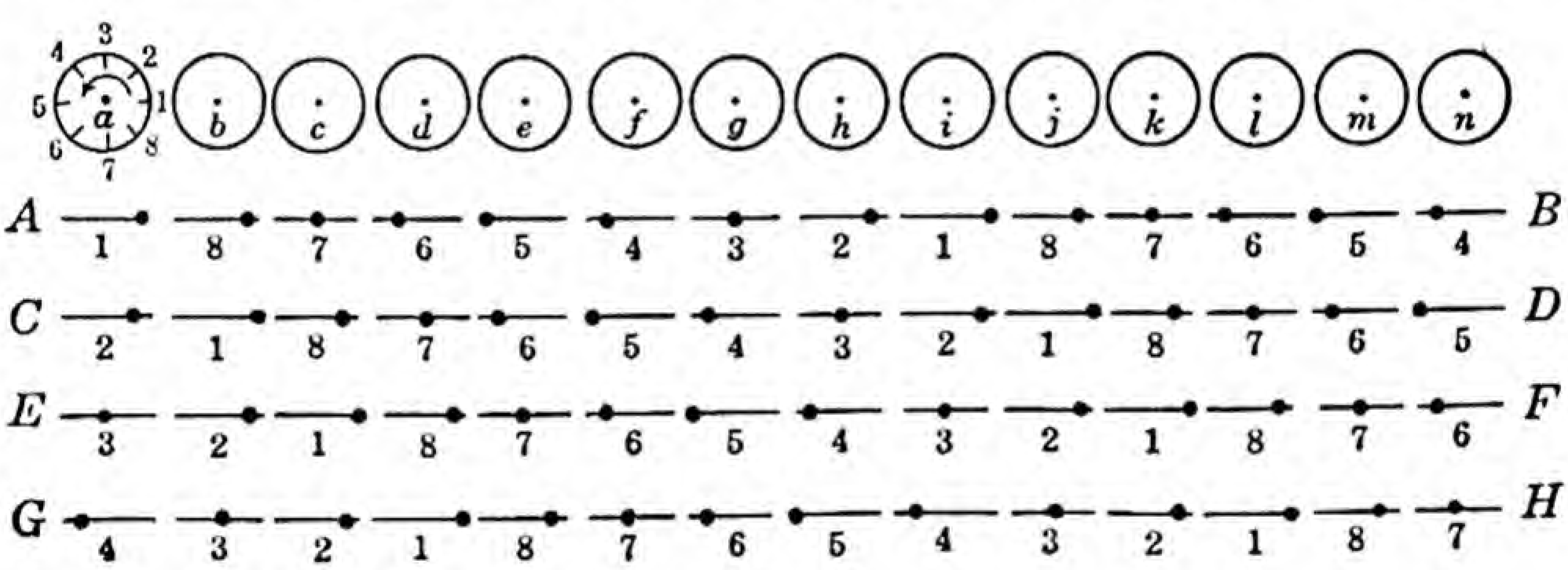


FIG. 53

along the direction of advance of the wave. In Fig. 53 are represented the relative displacements of a set of particles selected to be 45° out of phase. The reference circles for studying the motion of each particle are drawn with their centers directly above the rest positions of the set of particles. When the train of compressional waves is moving toward the right, the positions of the particles at one given instant are shown on line *AB*. Particle *a* is displaced to the right corresponding to point 1 on the circle. Particle *b* is displaced to the right corresponding to point 8 on the circle because it must be behind in phase with respect to particle *a*. Particle *c* is at its rest position, point 7; particle *d* displaced to the left, point 6; etc. After a lapse of time of one eighth of a period of the motion of the particles, the relative positions are shown on line *CD*. The

phase of particle a has advanced 45° and is represented by point 2, particle b by point 1; etc. On the two succeeding lines of the figure are pictured the positions of the particles at successive eighths of a period. A study of the relative positions of the particles will show on line AB a compression with its center at particle c and another at particle k. A rarefaction has its center at particle g. An eighth of a period later, line CD, the compressions have advanced an eighth of a wave-length to particles d and l and the rarefaction to particle h, and like advances are shown in lines EF and GH.

It is to be noted that if we wish to represent on co-ordinate axes the displacements of particles in case of a compressional wave, we may represent the displacements to the right as distances on the positive y axis and displacements to the left by distances on the negative y axis. Then there will result a curve identical in form to that of Fig. 52. Compressions will appear on such a diagram where the curve cuts the axis with a negative slope and the rarefactions at points where the curve cuts the axis with positive slope.

Since the motion of the particles in either a longitudinal or a transverse wave may be represented by a sine curve, as in Fig. 52, which progresses along the x axis as time progresses, we may obtain a single analytical expression which will represent both such moving wave forms.

56. The Analytical Formula for a Progressive Wave. — Let y_0 be the amplitude of a set of particles. Since the motion of any given particle is simple harmonic, we may write

$$y = y_0 \sin \frac{2\pi}{T} t, \tag{95}$$

which gives the value of the displacement y of a particle at any time t after it has passed through its equilibrium position with a positive velocity. Now consider two particles such as g (Fig. 52), and another particle, g', a short distance to the right of it. The phase of this second particle is behind that of g. If t' is the time which will be required before g' has the same phase as g, then we may write for the particle g',

$$y = y_0 \sin \frac{2\pi}{T} (t - t'). \tag{96}$$

A study of the progress of the wave form in Fig. 52 will show that when g' has the same phase which g formerly had, the wave form

has advanced the distance between g and g'. Let x be the distance from the given particle g to the particle g', v be the velocity of travel of the wave form, λ be the wave-length, and T the period of vibration of each particle. Then $x = vt'$ and $\lambda = vT$ so that

$$\frac{t'}{T} = \frac{x}{\lambda}. \tag{97}$$

Then the equation of the moving wave becomes

$$y = y_0 \sin \frac{2\pi}{T}\left(t - \frac{x}{v}\right),$$

or

$$y = y_0 \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right). \tag{98}$$

This equation represents the displacement (longitudinal or transverse) in a moving wave at any distance x from a chosen origin at any time t after the particle at the origin passes through its position of equilibrium with a positive velocity.

57. General Wave Equations. — It is seen in Eq. 98 that y is a function of two variables, x and t. If we consider t as a constant, we may take the derivative of y with respect to x. In giving t any chosen value, we effectively take a snapshot of the wave form at that value of t. The wave form for two such fixed values is shown in Fig. 52 by the heavy and dotted lines. Consider the displacements of two particles, one on each side of a particle such as g. Let the displacements of these two particles differ by an amount Δy and let their distance apart be Δx. As particles closer and closer to g are taken, the ratio $\Delta y/\Delta x$ approaches some limiting value called the partial derivative of y with respect to x at the point g. The partial derivative is indicated by the symbol $\partial y/\partial x$.

If the displacements (whether transverse or longitudinal) are plotted vertically as in Fig. 52, then $\partial y/\partial x$ gives the tangent to the curve. Physically, however, $\partial y/\partial x$ indicates the space rate at which the displacement varies for any given value of t. Considering t as constant and differentiating with respect to x, we obtain

$$\frac{\partial y}{\partial x} = -\frac{2\pi y_0}{Tv} \cos \frac{2\pi}{T}\left(t - \frac{x}{v}\right).$$

Now consider x as a constant and find the partial derivative of y with respect to t.

$$\frac{\partial y}{\partial t} = \frac{2\pi y_0}{T} \cos \frac{2\pi}{T}\left(t - \frac{x}{v}\right).$$

This expression gives the velocity of any particle in its path of vibration. In Fig. 52 it gives the velocity in the vertical direction. In Fig. 53 it gives the velocity of each particle along the line AB.

Comparing the expressions for the two partial derivatives, we see that

$$-v\frac{\partial y}{\partial x} = \frac{\partial y}{\partial t}. \tag{99}$$

We thus see that the two partial derivatives are always proportional to each other. The proportionality factor is the velocity of advance of the wave form. When

$$\frac{\partial y}{\partial t} = 0, \text{ then } \frac{\partial y}{\partial x} = 0,$$

i.e. when the velocity of a particle is zero (at maximum displacement) neighboring particles have the same displacement. When $\partial y/\partial t$ is a maximum, $\partial y/\partial x$ is a maximum, — *i.e.* when a particle is passing through its position of equilibrium, the difference in displacements of neighboring particles is a maximum.

Another very important relation is found by studying the relation of the two second partial derivatives:

$$\frac{\partial^2 y}{\partial x^2} = -\frac{4\pi^2 y_0}{T^2 v^2}\sin\frac{2\pi}{T}\left(t - \frac{x}{v}\right),$$

and

$$\frac{\partial^2 y}{\partial t^2} = -\frac{4\pi^2 y_0}{T^2}\sin\frac{2\pi}{T}\left(t - \frac{x}{v}\right).$$

From these equations we see that

$$v^2\frac{\partial^2 y}{\partial x^2} = \frac{\partial^2 y}{\partial t^2}. \tag{100}$$

In the next two sections important applications will be made of Eq. 100.

58. The Velocity of a Transverse Wave on a String or Flexible Wire. — Let ab (Fig. 54) be a portion of a string or wire along which a wave is passing. Consider a very small arc $AB = s$ whose center of curvature is at O. The tension T in the string acts tangentially along the arc. At A the tension acts at an angle which is different from that at which it acts at B. As shown in the vector diagram, the resultant of these two tensions

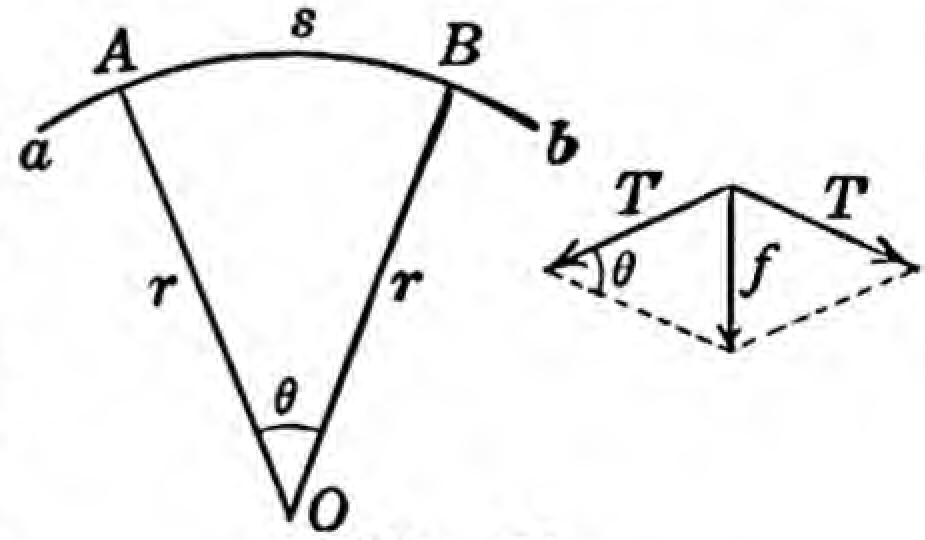


FIG. 54

is a force f which acts perpendicularly on the arc s. The vector triangle is similar to triangle ABO, so, replacing the chord AB by its arc, we have

$$\frac{T}{f} = \frac{r}{s}. \tag{101}$$

The radius of curvature of any portion of the curve is obtained from the expression

$$r = \frac{\left[1 + \left(\frac{dy}{dx}\right)^2\right]^{\frac{3}{2}}}{\frac{d^2y}{dx^2}}.$$

If the waves along the string have small amplitudes, the slope dy/dx is always small compared to unity and therefore $(dy/dx)^2$ is entirely negligible compared to unity. We may then write $1/r = d^2y/dx^2$. If we let m represent the mass per unit length of the string, we may apply Newton's second law to the mass ms and write

$$f = ms\frac{\partial^2 y}{\partial t^2} = \frac{Ts}{r} = Ts\frac{\partial^2 y}{\partial x^2}$$

or

$$\frac{T}{m}\frac{\partial^2 y}{\partial x^2} = \frac{\partial^2 y}{\partial t^2}. \tag{102}$$

Here also, since y varies with both x and t, the rate of change of y with respect to either x or t must be written as a partial derivative.

Comparing Eq. 102 with Eq. 100, we get for the velocity of a comparatively flat wave on a string,

$$v = \sqrt{\frac{\text{Tension}}{\text{Mass per unit length}}}. \tag{103}$$

59. The Velocity of a Compressional Wave. — As explained in the first section of this chapter, the velocity of propagation of a wave in an elastic body will depend on its modulus of elasticity and its density. The exact relationship will now be derived for a compressional wave.

Let us consider a series of waves with plane wave fronts. In Fig. 55, let the x direction be the direction of advance of the waves. As before, y represents the magnitude of the displacement of material along the x axis. This displacement varies along the x axis as shown in Fig. 53. The points c and d in Fig. 55 represent the situation at the boundaries of a very small region of length dx along the x axis. At C the compression carries the whole plane

of molecules a distance y over to point c. At a point D distant dx from C, where the phase of the motion is not as far advanced as at C, the plane of molecules is displaced from D to d a distance y'. The displacement y' is less than y so the material is compressed. Since C and D are only an infinitesimal distance dx apart, the difference in y and y' must be infinitesimal, which we indicate as dy. Not only is the material compressed, but it is slightly displaced as a whole. Therefore we may apply both Newton's second law and Hooke's law.

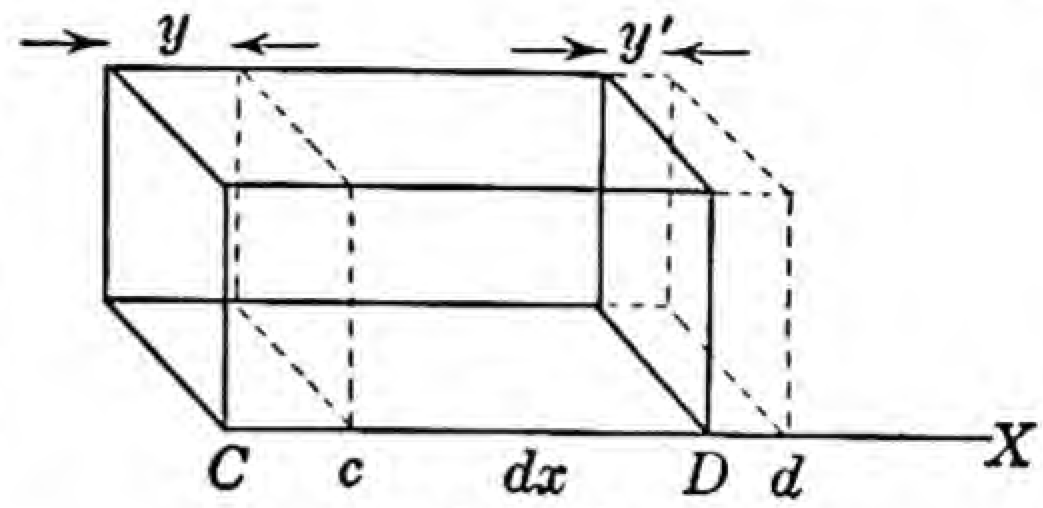


FIG. 55

If ρ is the density of the material, the infinitesimal mass included in a figure of cross section A and length dx is $\rho A\,dx$. The resultant force on this mass is the difference between the applied force at C which is directed to the right and the elastic return force at D directed to the left. Call this resultant force dF. It is to be noted that F and y each vary with respect to both the displacement x and the time t. Using the notation of the partial derivatives, we apply Newton's second law and obtain

$$dF = \rho A\,dx\,\frac{\partial^2 y}{\partial t^2},$$

or

$$\frac{\partial F}{\partial x} = \frac{\rho A\,\partial^2 y}{\partial t^2}. \tag{104}$$

By definition of the bulk modulus,

$$\beta = \frac{F}{A} \div \frac{\partial y}{\partial x}.$$

Solving this expression for F and differentiating it with respect to x we obtain

$$\frac{\partial F}{\partial x} = A\beta\,\frac{\partial^2 y}{\partial x^2}. \tag{105}$$

From Eqs. 104 and 105, we have

$$\frac{\beta}{\rho}\frac{\partial^2 y}{\partial x^2} = \frac{\partial^2 y}{\partial t^2}. \tag{106}$$

Comparing this last equation with Eq. 100, we see that

$$v = \sqrt{\frac{\beta}{\rho}}. \tag{107}$$

In § 13, we proved from the theory of dimensions that the equation was of the form of Eq. 107, but then we were not able to prove that the proportionality constant k was unity.

60. The Velocity of a Compressional Wave in a Perfect Gas. — We must realize that the modulus of elasticity of a substance may be measured under many conditions of temperature and heat content. The two important values are the isothermal and adiabatic moduli. The ratio of the stress to the strain when the temperature is maintained constant is called the isothermal modulus of elasticity. When no heat is allowed to flow in or out of the material, the temperature will rise or fall and the ratio of the stress to the strain is called the adiabatic modulus of elasticity. If the pressure is increased suddenly, the temperature of a gas rises and the change in volume is not as large as if the gas were allowed to cool down to the original temperature before the change in volume was measured. In § 84, it is shown that the isothermal modulus of elasticity of a perfect gas is numerically equal to the existing pressure of the gas, while the adiabatic modulus of elasticity is k times as large, k being the ratio of the two specific heats of the gas. So, for a perfect gas, we deduce theoretically that

$$v = \sqrt{\frac{kp}{\rho}}. \tag{108}$$

Experimentally, the value of k for air at 0° C. is found to be 1.403, so for air at 0° C. and standard pressure, we may compute the velocity of sound as follows:

$$v = \sqrt{\frac{kp}{\rho}} = \sqrt{\frac{1.403 \times (76 \times 13.60 \times 980)}{0.001293}}$$
$$= 331.5 \text{ meters per second.}$$

The velocity of sound at 0° C. in air is found experimentally to be about 330.4 meters per second or 1084 feet per second.

We know that the pressure of the air, just as for all fluids, is caused by the weight of the air above the earth. An increase or decrease in temperature, which causes the molecules to change their distance apart, could not produce any variation of the gravitational force; therefore, a variation in temperature will not alter the value of p in Eq. 108. Nevertheless the change in temperature will affect the value of ρ and thus change the velocity of sound.

If the pressure of the gas is changed, keeping the temperature constant, we predict by Boyle's Law that the density of the gas will

increase in the same proportion as the pressure and so the velocity of sound in a gas should be independent of pressure changes.

The effects of both pressure and temperature as predicted by Eq. 108 are found to be in strict agreement with experiment.

61. Interference between Two Wave Trains. Standing Waves. — By a study of Fig. 53 and the analytic geometry of the equations, the student is expected to show that just as the equation

$$y_1 = y_0 \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right) \tag{96}$$

represents a wave traveling in the positive x direction, the equation

$$y_2 = y_0 \sin 2\pi\left(\frac{t}{T} + \frac{x}{\lambda}\right) \tag{109}$$

represents a wave of the same frequency, wave-length, and velocity traveling in the opposite direction.

When these two wave trains are passing in the same straight line through a medium, each train produces its own displacement and the resulting combined displacement is represented by the sum of the two separate displacements. Thus

$$y = y_1 + y_2 = y_0\left[\sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right) + \sin 2\pi\left(\frac{t}{T} + \frac{x}{\lambda}\right)\right],$$

which reduces to

$$y = 2\,y_0 \sin 2\pi\left(\frac{t}{T}\right) \cos 2\pi\left(\frac{x}{\lambda}\right). \tag{110}$$

The way in which this resultant displacement varies with position and time will now be analyzed.

For any given value of x, the value of the displacement is seen to vary as a sine function of the time. The amplitude of the displacement is

$$2\,y_0 \cos 2\pi\left(\frac{x}{\lambda}\right).$$

It is easily seen that, for values of x equal to any odd multiple of $\lambda/4$, the amplitude is zero at all times, while for all even multiples of $\lambda/4$, there is simple harmonic motion with a maximum amplitude of $2\,y_0$, the sum of the amplitudes of the two separate waves. These points of zero and maximum amplitudes are called "nodes" and "loops" respectively. The range of the amplitudes at all values of x is shown in Fig. 56 by the values of y between the heavy and light full-line curves.

The above analysis has shown only the maximum value of the displacements. We must see how the amplitudes of the various particles are related to each other at given instants. From Eq. 110 we see that when $t = 0$, $y = 0$ so that the displacements are represented by the x axis. The student should show that Eq. 110 is represented graphically

by the dotted curve when $t = \frac{T}{12}$,
by the heavy full-line curve when $t = \frac{T}{4}$,
by the dotted curve when $t = \frac{5}{12} T$,
by the x axis when $t = \frac{T}{2}$,
by the dashed curve when $t = \frac{7}{12} T$,
by the light full-line curve when $t = \frac{3}{4} T$,
by the dashed curve when $t = \frac{11}{12} T$,
and by the x axis when $t = T$.

These combinations of waves are appropriately called *standing waves.* Fig. 56 shows the actual appearance of a set of standing waves produced by two trains of transverse waves on a string. In

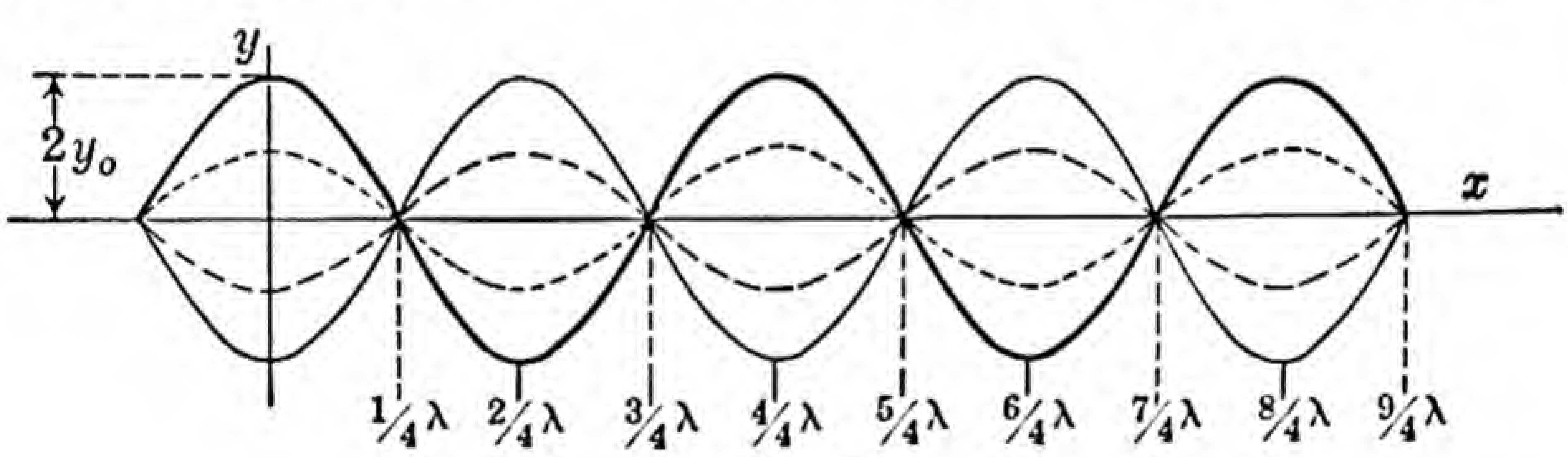


FIG. 56

the case of a set of compressional standing waves as in an organ pipe, the values of the ordinates of the heavy curve indicate the maximum displacement which the particles undergo in their vibration back and forth along the length of the pipe.

62. Change of Phase on Reflection. — One of the simplest means of producing a reversed train of waves of the same frequency as the direct train of waves, is to have the wave reflected so as to travel back along the same path as the incident train.

It is easy to observe a single transverse wave travel down a stretched cord and travel back after reflection at the fixed end. Careful observation will show that during the reflection a change in phase of 180° will occur. We may predict this result by the

following argument. At any node the two waves must always be 180° out of phase or else the resultant displacement could not be zero continuously. The fixed end of a stretched cord is observed to be a point of zero displacement. Therefore the reflected waves must be 180° out of phase with the incident waves. This state of affairs must exist at the end of a closed organ pipe. The molecules immediately against the end wall are motionless, therefore the wall represents a place where the displacements of the separate direct and reflected waves are at all times equal and opposite in sign or, in other words, 180° out of phase. To find what a compression or rarefaction becomes after reflection, we may refer to the graphical representation of the direct and reflected components of the standing waves as shown in Fig. 57. In part (*a*) the heavy dotted line represents the displacement of particles for a train of waves traveling to the right, and the light dotted line the displacement for a reflected train moving to the left. The reflecting wall must be represented at such a place as the line *AB*, or any similar place where the displacements are equal and opposite. Of course, the waves to the right of *AB* do not exist but are drawn only as continuation of the other lines. In part (*a*) the direct train has a compression *C* approaching the wall (see § 55). The reflected wave has a rarefaction slightly to the left of *C* and a phantom compression *c* which is approaching as shown by the imaginary portion of the wave. The combined effect of these two trains, the resultant actual wave, is shown by the full curve. The state of affairs a moment later is shown in part (*b*). The two trains coincide, each producing equal displacements at the same place, and the resultant displacement is shown by the full curve of double the value due to each single train. A short time later, conditions exist as shown in part (*c*). The compression *C* is shown as a phantom wave behind the wall,

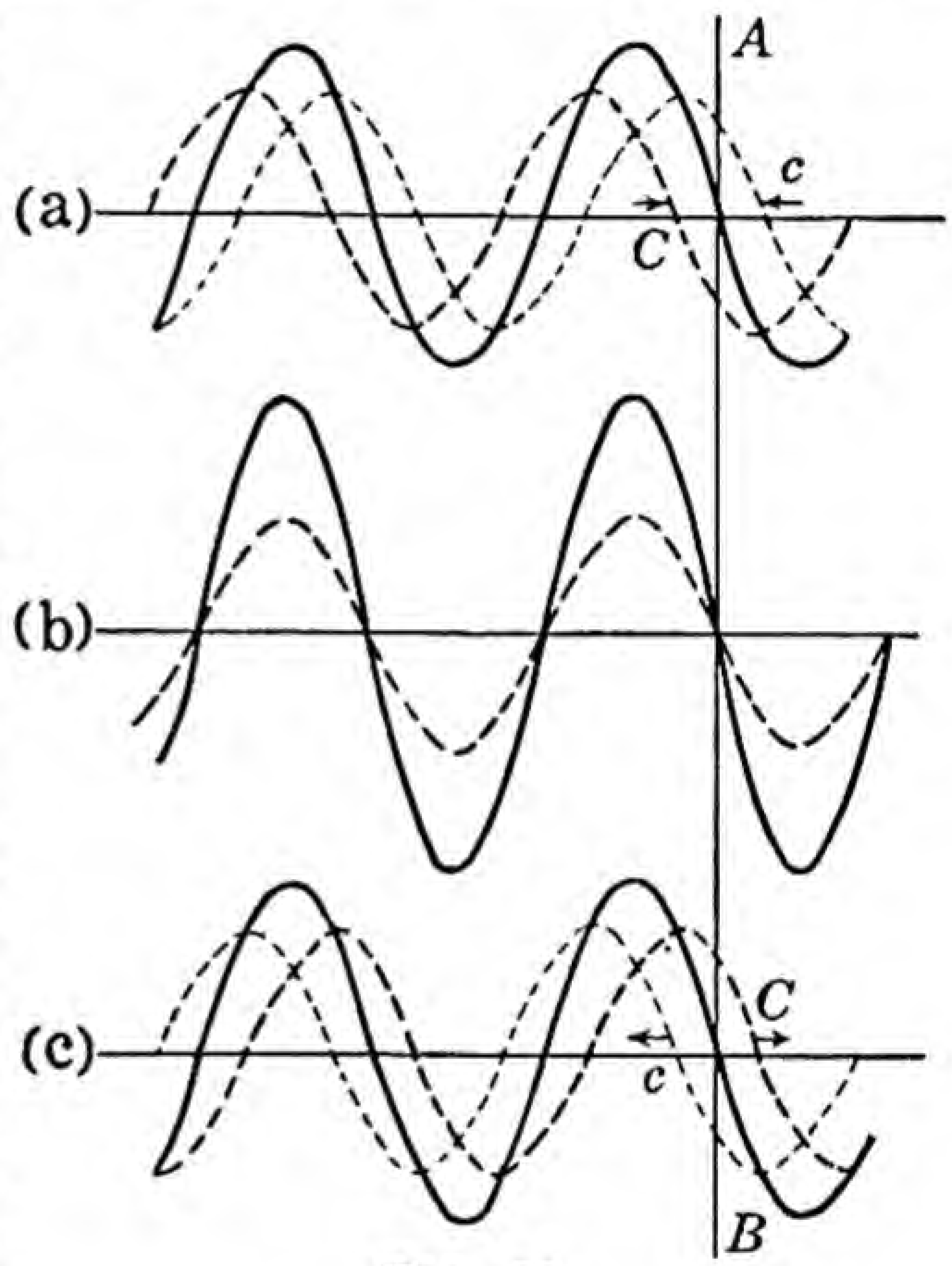


FIG. 57

while the compression c has emerged from the wall and is traveling away from it. The result clearly shows that a compression such as C is reflected as a compression such as c and a rarefaction is reflected as a rarefaction. This is the result only when the reflection occurs with 180° phase difference. (At the wall the displacements produced by the separate waves are equal and opposite.) The full curves show the progress of the resulting standing waves.

In a like manner we could show that when a wave reaches the end of an open organ pipe, where there is a loop, the reflection occurs in phase, and as a result we find that a compression is reflected as a rarefaction and a rarefaction as a compression.

We observe that in Newton's rings (see § 223) the center ring is black and also that the thinnest part of a soap bubble just before it bursts turns black. We know that blackness is caused by destructive interference between the light reflected from the front air-to-glass surface and that reflected from the back glass-to-air surface. The distance between the paths of these waves is not sufficiently long to cause one group of waves to be appreciably out of phase with the other, so we conclude, from the case of the open and closed organ pipes, that at one of the surfaces the reflection takes place with a phase 180° different from that which occurs at the other surface.

63. The Doppler Effect. — In 1842, an Austrian physicist, Doppler, announced that light from the stars should be affected by the motion of the stars. Doppler had previously shown the corresponding effec in sound which we shall describe. When a source of sound having a definite pitch is moved toward an observer the observed pitch is higher than when the source is at rest. A lowering of the pitch results when the source recedes from the observer. There is likewise a change in pitch caused by the motion of the observer and in certain cases by the motion of the medium through which the waves are passing.

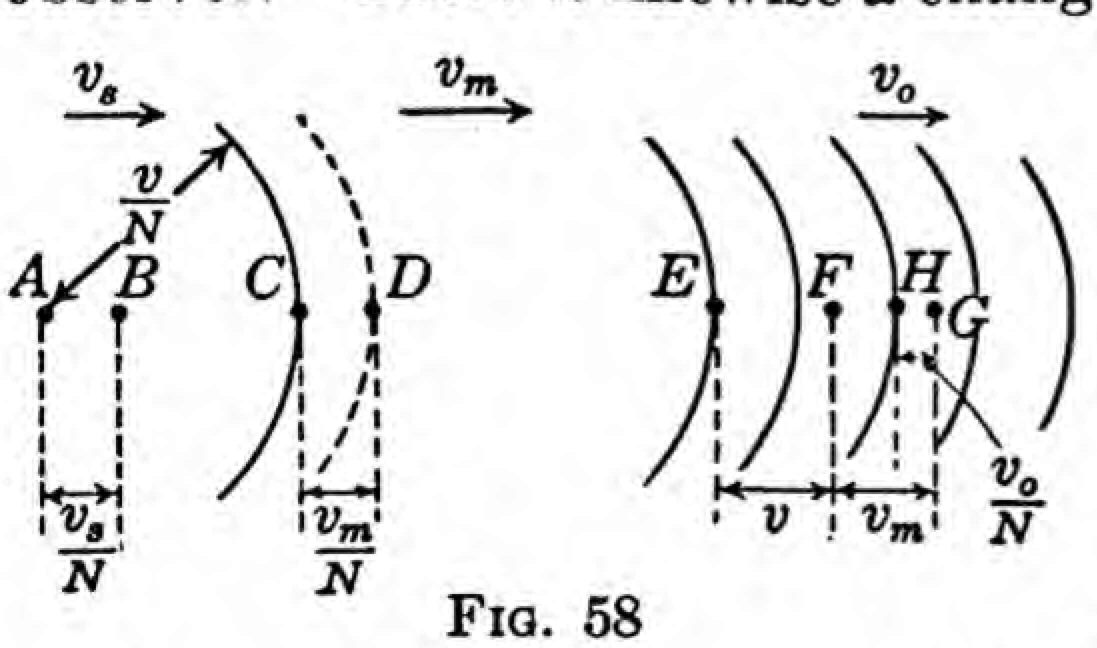


FIG. 58

Consider the general case where the source, observer, and medium are all moving. The source at A (Fig. 58) is emitting waves of frequency N and these waves "travel" through the medium with a velocity v. The source is moving to the right

with a velocity v_s. A wind is blowing the medium toward the right with a velocity v_m and the observer is moving in the same direction with velocity v_o.

(*a*) *The Effect at the Source.* — Let the source be at A when it is in any given phase of its vibration. After $1/N$ part of a second it is in the same phase of its vibration. The distance between the source at the end of that time and the position of the wave front emitted at the beginning of that time will be the actual wave-length. Due to the elastic properties of the medium, the wave front would have traveled the distance $AC = v/N$, the wind will carry it a distance $CD = v_m/N$ during that time, and the source will have moved a distance $AB = v_s/N$. So the distance BD between successive waves will be the wave-length λ. Thus

$$\lambda = \frac{v}{N} + \frac{v_m}{N} - \frac{v_s}{N} = \frac{v + v_m - v_s}{N}. \tag{111}$$

If the wind velocity is constant, there is nothing to disturb the continuous passage of waves of this wave-length across the whole distance from the source to the observer.

(*b*) *The Effect on the Observer.* — The problem now becomes: How many vibrations per second does the observer's ear receive? If the observer were stationary and there were no motion of the medium, he would receive each second all waves which were between him and a point $EF = v$ cm. nearer the source. That number of waves would be v/λ. But due to the fact that the waves are carried with their medium a distance $FG = v_m$ farther during that second, the observer will receive v_m/λ more vibrations per second. Finally due to his motion away from the source with a velocity v_o, the observer will fail to receive during that second v_o/λ waves, those located over a distance $HG = v_o/N$. He will, therefore, receive

$$N' = \frac{v + v_m - v_o}{\lambda} \text{ vibrations per second.} \tag{112}$$

The actual number of vibrations emitted per second from the source, which determines the pitch of the sound with everything at rest, is N. Due to the motion, the apparent pitch is N'.

$$\frac{N'}{N} = \frac{v + v_m - v_o}{v + v_m - v_s}. \tag{113}$$

It is to be noted that v_s is positive whenever the source is moving toward the observer, v_o is positive when the observer is moving

away from the source, and v_m is positive when the medium is moving from the source to the observer. Whenever any of the velocities is reversed, the sign of that velocity must be reversed in Eq. 113.

By examining the light from certain stars with a spectroscope, the lines of the spectra due to known chemical elements are found to be shifted slightly from their positions as determined from stationary laboratory sources. Only a change in wave-length can account for this shift and only the motion of the star can account for the change in wave-length. When light from opposite sides of a rotating planet is observed, the shifts of the lines are in opposite directions. After correction for the motion of the earth is made, the velocity of rotation of the planet is found. Much work is now being done in computing the velocities of distant stars from the observed Doppler shift of their spectral lines.

PROBLEMS

1. An automobile traveling at the speed of 60 miles per hour has a horn which has a pitch of 500 vibrations per second. What is the apparent pitch of the horn to a person standing still as the car approaches? What change of pitch occurs as the car passes by? Answer the same questions for an observer in another car traveling in the opposite direction with the same speed.

2. Show that if the observer and source are both stationary, then the observer will hear the correct pitch even though a steady wind is blowing.

3. If the observer is stationary in motionless air, what would be the speed required in order that the pitch of a passing source drop a whole octave? A half tone (see § 65)? A whole tone ($\frac{9}{8}$)? Answer for the case of the stationary source and a moving observer.

64. The Physical Basis of Music. Harmonics. — A vibrating system such as a violin string, tuning fork, organ pipe, etc., is a system in which standing waves are being maintained, these waves being due to the superposition of waves of equal amplitude, traveling in opposite directions. The points of constraint, such as the ends of the string, the base of the tuning fork,* etc., are necessarily nodes. In an open organ pipe the end is a point of maximum motion or a loop; in a closed pipe the closed end is a node. There may be nodes between the points of constraint depending

* The middle of the base of a tuning fork is actually the center of a short loop, the nodes being near the base of the prongs. The increased mass in the base of the fork results in a short segment in this region, a segment only a few centimeters in length; hence, the tuning fork is not exactly the same as a uniform rod bent into a U-shape.

upon the frequency of the wave, but there must be an integral number of vibrating segments between the points of constraint. The system can maintain standing waves or support vibrations of only those frequencies for which this condition is fulfilled.

(*a*) *Constraint at Both Ends.*—In case of a vibrating string or any vibrating linear system in which no motion is allowed to take place at its ends, the lowest possible frequency of vibration is had when there is a loop midway between the two ends. The length of the string is thus one half of a wave-length of the waves traveling along the string. The system may be made to vibrate in two sections with a node at the center or in many sections with *any number of nodes* evenly distributed between the ends.

(*b*) *Constraint at Center.* — A vibrating system such as a rod clamped at the center must always have a node at its center. The lowest frequency is obtained when there is a loop at each end. Higher frequencies are obtained when *any odd number of nodes* exist.

(*c*) *Constraint at One End.* — The simplest mode of vibration of a linear system in which a node is maintained at one end is obtained when a loop exists at the other end. The next higher frequency exists when a node appears at a point two thirds the length of the system from the motionless end. Still higher frequencies exist when a greater number of nodes exist.

The student is expected to show in which of the above cases the possible higher frequencies are only the odd multiples of the lowest frequency and in which cases all multiples may exist.

The lowest frequency with which a system may vibrate is called the *fundamental.* Frequencies which are multiples of the fundamental are called *harmonics.* Due to peculiar constraints, such as loading unsymmetrically, frequencies which are not multiples of the fundamental may exist. All the frequencies existing on a vibrating system are called *overtones.* Only those which are multiples of the fundamental are called harmonics. The quality of a musical note depends upon the number and relative intensity of the overtones. The quality may be altered by the use of resonance chambers, sounding boards, etc., which respond strongly to certain overtones. The blending of two tones one of which is a harmonic of the other is always pleasing to the ear.

65. The Diatonic Scale. — Our musical scale is made up of a sequence of three notes whose frequencies are in the ratios 4 : 5 : 6. This combination of notes has many overtones in common and

produces no beats between any of the strong overtones. That selection was no doubt made because of the pleasing sound made when the three tones are produced at the same time. The notes form the C, E, and G of our scale (do, mi, sol) and form the major chord. For the pitch of the C at about the middle of the pianoforte (middle C), 264 vibrations per second is usually selected. This results in 330 vibrations per second for E and 396 for G. Now if G is taken as the starting point of a new major chord, we obtain $4:5:6::396:495:594$. Again if the octave of middle C, 528, be taken as the highest note of a major chord we obtain

$$6:5:4::528:440:352.$$

Now if we use the octave below 594, namely 297, and arrange all of these notes in the order of frequency, we obtain the succession of notes which we call the diatonic scale. (See the highest three lines of the following table for these separate notes in their proper order and the fourth line for the complete sequence including the octaves of several notes.) This succession of notes was gradually developed and selected by Europeans as being a very harmonious group of notes and has now become very widespread over the world.

Many attempts have been made to add more notes to an octave. Probably the only one in use now is a "quarter-tone" scale with one note about half way in pitch between each two successive notes in our thirteen-note octave. At present some of the tonal combinations with such a scale are pleasing, some rather disturbing, and many actually unpleasant. In time, we could get accustomed to such a scale, no doubt.

The ratio of the frequencies of each note to the note lower in pitch is indicated below.

$$\begin{array}{ccccccccccccccc} C & & D & & E & & F & & G & & A & & B & & C \\ & \frac{9}{8} & & \frac{10}{9} & & \frac{16}{15} & & \frac{9}{8} & & \frac{10}{9} & & \frac{9}{8} & & \frac{16}{15} & \end{array}$$

The ratio $\frac{9}{8}$ and $\frac{10}{9}$ are not far different from each other, while the ratio $\frac{16}{15}$ which occurs twice in the scale represents about one half as large an interval in pitch. For that reason the steps from E to F and B to C are called half tones while all other steps are called whole tones.

The first instruments made had only the above succession of notes and their octaves. If a certain tune started with C, say, then it was found that the same tune could not be played on that

Key Note	C	D	E	F	G	A	B	C	D	E	F	G	A
C	264		330		396 396		495		594				
	264	297	330	352 352	396	440 440	495	528 528	594	660	704	792	880
G					396	445.5	495	528	594	660	F # 742.5	792	
D		297	334.1	371.3	396	445.5	495	C # 556.9	594	668.3	742.5	792	
A					412.5	440	495	550	586.7	660	733.3	G # 825	880
E			330	371.3	412.5	440	495	550	D # 618.8	660	742.5	825	
B					412.5	464.1	495	556.9	618.8	660	742.5	825	A # 928.1
F #													
F				352	396	440	B ♭ 469.3	528	586.7	660	704	792	
B ♭					391.1	440	469.3	528	586.7	E ♭ 625.8	704	782.2	
E ♭					391.1	A ♭ 417.2	469.3	521.5	586.7	625.8	704	782.2	
A ♭					391.1	417.2	469.3	521.5	D ♭ 556.2	625.8	695.3	782.2	
D ♭					G ♭ 370.8	417.2	463.5	521.5	556.2	625.8	695.3	741.7	
G ♭													
Tempered Scale				370.0	415.3	466.2		554.4	622.3		740.0	830.6	
			329.1	349.2	392.0	440	493.9	523.3	587.3	659.3	698.5	784.0	880

instrument if it were started on any other note. There would be needed certain tones which would be found to lie part way between some of the notes where a whole step occurred. Those desired notes can be found as follows. Let a new diatonic scale be started on G, 396 (line 5 of table). Taking the ratio $\frac{9}{8}$, the frequency of the next note needed in the scale is 445.5; then follow 495, 528, 594, 660, and 742.5. Two new frequencies appear: 445.5 and 742.5. The number 445.5 is quite close to 440; in fact so close that either note would sound well to the ear when used in either scale. But the note 742.5 is too nearly in the middle of the whole step from F to G. This new note is called F sharp (F♯). This scale is said to be in the key of G and has one sharp which is inserted as a black note on the piano. In a like manner if a scale is started on D (line 6 of table), a note 664 could be used in place of 660 and 668.25, but an entirely new note 556.9 called C♯ is needed. Likewise the key of A needs the additional G♯, the key of E needs D♯, and the key of B needs A♯. Thus we call the key of G the key of one sharp, the key D the key of two sharps, etc. The playing on a piano of the scale in the key of D demands the use of two black notes, the key of A three black notes, etc. The student is expected to work out the frequencies for F♯, the key of 6 sharps, and to name the sixth sharp.

It is seen that we have not yet started a key of F or any of the five new half tone notes. Starting on F, 352, we need an extra tone of frequency $469\frac{1}{3}$ — called B flat (B♭). This so nearly agrees with A♯ that a compromise note of 467 would not be disagreeable. In the key of B♭ a note E♭, 625.8, is needed, but D♯ may be modified to suit. Then the key of E♭, A♭, D♭, and G♭ may be computed, needing A♭, D♭, G♭, and C♭, respectively.

66. The Tempered Scale. — If a piano were built so that a correct diatonic scale could be played, starting on any given note, about thirty separate notes would be required for each octave. Many of the notes would be very close in pitch to neighboring notes, — so close that only a person with especially sensitive ears would notice the difference. With these "tempered" notes introduced, only 12 notes are needed in each octave and such a tuned set of 12 notes constitutes the "tempered scale."

The piano tuner ordinarily tunes middle C so that no beats occur between it and a standard tuning fork. Then the major chord is tuned (C, E, and G) until unpleasant beat-notes disappear,

and other combinations of the sequence 4, 5, 6 similarly tuned. Going from one group of notes to another, the first octave is tuned and tempered. Then the other notes are tuned in octaves from the notes of the original octave and then tested to ascertain temperament. Evidently a good piano tuner must have an ear very sensitive to the presence of beat-notes and to various discordant combinations of notes in order to judge when the scale is well tempered, — *i.e.* sounding about equally bad (or good) in all keys. Obviously a violin player may play in the true or "just" scale if he has sufficient accuracy of touch and sufficiently fine sense of pitch to detect the faults in a tempered scale.

A scale of twelve notes per octave is said to be equally tempered when the ratio of the frequency of any note to that of the preceding note is equal to the twelfth root of two. Thus in the last line of the table,

$$(440)2^{\frac{1}{12}} = 466.2, \ (440)2^{\frac{2}{12}} = 493.9,$$

etc. The frequency of each note of the equally tempered scale differs only slightly from the average value of the corresponding frequencies needed in the separate keys.

The A above middle C is taken as the standard pitch for orchestral instruments. Its actual pitch has varied through the last century from 435 to 460 vibrations per second. In 1925, musicians and manufacturers of musical instruments in the United States selected A440 as the standard musical pitch for this country.

CHAPTER III

HEAT

Thermometry

67. Temperature Scales. — Our first notions of temperature are obtained through our physiological sensations. We say that a body is at a "high" or a "low" temperature according to the way it feels. Observation shows that the various physical properties of materials, such as the length of a solid, the volume of a liquid, the pressure and volume of a gas, the electrical resistance of a metal, etc., change on heating. These changes have in various ways been adopted as measures of temperature. It has been customary to select some property arbitrarily and in terms of this property to define a temperature scale. The change of volume of mercury in glass affords a most convenient measure of temperature, and scales chosen in terms of this property are known to everyone.

A substance whose properties may be used as a measure of temperature is called a *thermometric substance*, the property chosen, a *thermometric property*. Any temperature measuring device is a thermometer.

The two fixed points on all temperature scales are the freezing point and boiling point of pure water under one standard atmosphere of pressure. The standard atmosphere of pressure is defined to be that which is exerted by a column of mercury 76 cm. high of normal density at 0° C. and under a gravitational attraction such that $g = 980.665$ cm./sec^2. or 32.1740 f ./sec^2. This is called the normal value of g at sea level at 45° latitude. The student should explain why the gravitational attraction must be considered. The Centigrade degree for any thermometric substance is the change of temperature which causes the thermometric property of that substance to change 1/100th part of its change between the ice and steam points, and the ice point is called 0° C. There are 180 intervals of the Fahrenheit scale between the steam and

ice points, and the ice point is called 32° F. It is to be carefully noted that the temperatures of 0° C. and 100° C. are perfectly definite, being fixed by the properties of water, but the Centigrade temperature of any other point will depend upon the properties of the particular substance chosen for the thermometer.

Let P_0 and P_{100} be the values of the thermometric property of the chosen substance at the fixed points and P_t the property at some other temperature t. The quantity $P_t - P_0$ is the change in the property when the temperature is raised from 0° C. to t° C. The value of this temperature t is defined to be the fractional part of 100° that $P_t - P_0$ is of the total change $P_{100} - P_0$. Therefore we have the defining equation

$$t = \frac{P_t - P_0}{P_{100} - P_0} 100. \qquad (114)$$

This is the general expression for temperature on any Centigrade scale.

Suppose that an iron bar is transferred from an ice bath to a bath whose temperature is such that the bar expands 25 per cent of its expansion between 0° C. and 100° C. Using the iron bar as an expansion thermometer, the temperature is 25° C. There is no reason to suppose that a mercury in glass thermometer or any other thermometer placed in the same bath at the same time with the iron rod would undergo 25 per cent of its respective expansion. To be sure, the differences in the readings on the different types of thermometers are small, but they are measurable and have to be taken into account for accurate work. So for all precise work, not only the readings of the scale, but the material of the thermometer must be stated.

If the change in any thermometric property is plotted against the temperature t (measured from the change in property of some thermometer taken as a temporary standard), the graph will in general not be a straight line. The curve may be accurately represented by a series,

$$P_t = P_0(1 + at + bt^2 + ct^3 + \cdots), \qquad (115)$$

where a, b, c, etc., are constants depending upon the substance and the temperature scale. Usually b is much smaller than a, and c much smaller than b, etc. See Problem 7, § 68, for a typical equation. In that problem the term at is quite significant when compared with the first term of unity. The term bt^2 is much less

important as long as t is small. The larger t becomes, the larger and thus more important do the higher terms of this series become. Eq. 115 may be represented by a curve such as the heavy line in Fig. 59. Consider the tangent to the curve at C, where $t = 0$.

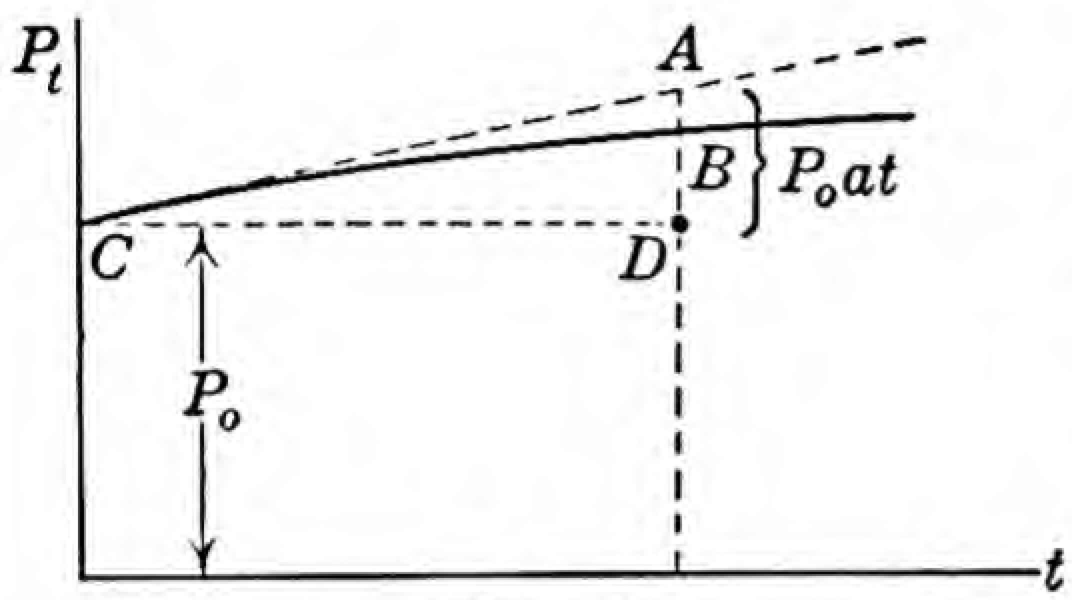


FIG. 59

Over the region CA, where AB is not an appreciable part of AD, the curve may be represented by the straight line CA, which is given by the first two terms of Eq. 115, namely:

$$P_t = P_0(1 + at). \quad (116)$$

The distance AB is equal to

$$P_0(bt^2 + ct^3 + \cdots),$$

and is small compared to AD if we do not follow out too far along the curve, *i.e.* if t is not taken too large.

68. The Expansion of Solids. — Let a bar of some solid be placed in successive baths of different temperatures. Let the length of the bar l_t be measured while it is in each bath and let the temperature of each bath be measured by any thermometer which is not of the same material as that of the bar. When applied to this case, Eq. 115 becomes

$$l_t = l_0(1 + at + bt^2 + ct^3 + \cdots),$$

where l_0 is the length of the bar at the ice point. Just as explained in § 67, if b, c, etc., are small and t is not too large, we may write

$$l_t = l_0(1 + at), \text{ or } a = \frac{l_t - l_0}{l_0 t}. \quad (117)$$

The value of a is called the *temperature coefficient of linear expansion* of the substance. It is the ratio of the change in length per degree change in temperature to the original length at the ice point, or it is the fractional part of its length at the ice point that a substance expands for each degree rise in temperature.

Very often the length l_1 of a bar is measured at some temperature t_1 and it is desired to know the amount of expansion caused by raising the temperature to t_2. Applying Eq. 117, we have

$$l_2 = l_0(1 + at_2), \text{ and } l_1 = l_0(1 + at_1).$$

Then
$$l_2 - l_1 = \Delta l = l_0 a(t_2 - t_1)$$
$$= \text{approximately } l_1 a(t_2 - t_1).$$

Due to the fact that a is very small for most solids, l_0, l_1, and l_2 differ from each other but little and therefore in numerical problems one may be substituted for the other wherever any one of them occurs as a *multiplying* factor.

Where the curvature of the line in Fig. 59 is considerable, a temperature coefficient at any temperature t_1 is defined as follows for any small range of temperature over which the curve is sensibly straight,

$$a = \frac{l_{t_2} - l_{t_1}}{l_{t_1}(t_2 - t_1)}.$$

QUESTIONS AND PROBLEMS

1. Name various thermometric properties of substances and describe how a thermometric scale might be based upon each.

2. State several reasons why a mercury-in-glass thermometer does not make a reliable standard.

3. Adapt Eq. 114 to the Fahrenheit scale.

4. Does the numerical value of the coefficient of expansion depend upon the units of length? Does it depend upon the temperature scale? Explain.

5. Show that the volume coefficient is approximately three times the coefficient of linear expansion. Interpret the approximations geometrically. Work out the corresponding case for area expansion of a rectangular plate.

6. A thermometer with an arbitrary scale of equal parts reads -40.0 in melting ice and 200.0 in water boiling at standard pressure. Find the Centigrade temperatures corresponding to 8.6 and 124.0 on the thermometer.

7. One junction of a certain thermocouple is kept at the ice point. The temperature of the other junction is read with a mercury-in-glass thermometer. The e.m.f. in micro-volts, as a function of the temperature of the thermometer, is given by the following equation:

$$\text{e.m.f.} = 0.05175\, t - 0.0000105\, t^2.$$

Using the e.m.f. as a thermometric property, what temperature will be indicated by the thermocouple when the mercury thermometer reads 50° C.?

8. A compensated pendulum consists of a very thin iron tube of length l, filled with mercury to a height h. If the mass of the tube is negligible compared with that of the mercury, show that, to a high degree of accuracy, $h = 2\, l\, a/(B - 2\, a)$, where a is the coefficient of linear expansion of the iron and B is the coefficient of volume expansion of mercury. If l is 108 cm., a is 12×10^{-6} per degree C., and B is 182×10^{-6} per degree C., show that h is 16.5 cm.

9. A bottle which is completely filled with milk contains one liter when pasteurized at 60° C. Taking the coefficient of cubical expansion of milk and glass to be 0.000380 and 0.000025 per degree Centigrade respectively, find the empty space in the bottle when the milk and bottle are cooled to the ice point.

10. The case around a certain mercury barometer is made of brass. The scale made on this case is correct at 0° C. Calculate the corrections necessary, due to the expansions of the mercury and brass, when the barometer is read at 25° C.

11. A glass vessel with a capillary opening is completely filled with mercury at 20° C. It is then raised to the steam point and some of the mercury runs out. From the following weighings, compute the coefficient of linear expansion of the glass.

Empty glass vessel	100.23 gms.
Vessel filled with Hg. at 20° C.	754.30 gms.
Vessel after cooling down from 100° C. . .	746.26 gms.

Expansion of Gases

The Fundamental Gas Laws and the Equation of State

69. The Equation of State. — When a gas is allowed to expand at constant temperature, it is found that the product of the pressure and the volume of the gas is approximately constant. This is known as Boyle's law. If the volume of a gas is kept constant and the pressure is used as the thermometric property, the temperature on such a scale is given by the expression

$$t = \frac{p_t - p_0}{\dfrac{p_{100} - p_0}{100}}. \tag{118}$$

We define the pressure coefficient of a gas, α, as the ratio of the change in pressure per degree to the pressure at the ice point, the volume remaining constant. To avoid any reference to a thermometer of a particular material, we choose 0° and 100° C. and write

$$\alpha = \frac{p_{100} - p_0}{100\, p_0}.$$

Combining this with Eq. 118,

$$t = \frac{p_t - p_0}{\alpha p_0} \text{ and } p_t = p_0(1 + \alpha t), \tag{119}$$

where t is the temperature on the constant volume gas pressure scale. The logic of this procedure should be noted. If the temperature t is measured on the gas pressure scale, then of course the change in the gas pressure is directly proportional to the temperature, by definition. The gas pressure plotted against the temperature must give a straight line. Since α is determined by the slope of the line and since the slope of a straight line is every-

where the same, α is the same whether computed from 0° C. to 100° C., as was done above, or from 0° C. to t° C., as $\alpha = (p_t - p_0)/p_0 t$. This latter expression is identical with Eq. 119.

If some other thermometric property had been chosen as a basis for measuring t, we should have had an expression similar to Eq. 115, $p_t = p_0(1 + at + bt^2 + ct^3 + \cdots)$, with Eq. 119 holding over a small range of temperature.

If the pressure of the gas is kept constant and its volume is used as the thermometric property, temperature on such a scale is given by the expression,

$$t' = \frac{v_{t'} - v_0}{\dfrac{v_{100} - v_0}{100}}. \tag{120}$$

Calling $(v_{100} - v_0)/100\, v_0 = \beta$, the volume coefficient, we have

$$t' = \frac{v_{t'} - v_0}{\beta v_0}, \text{ and } v_{t'} = v_0(1 + \beta t'). \tag{121}$$

The student should put into words the definition of β, the volume coefficient of expansion of a gas.

We now propose to show that for a perfect gas these two temperature scales are identical. Consider a certain amount of a gas enclosed in the bulb of a gas thermometer. Such a thermometer consists of a bulb connected to a U-tube containing mercury.

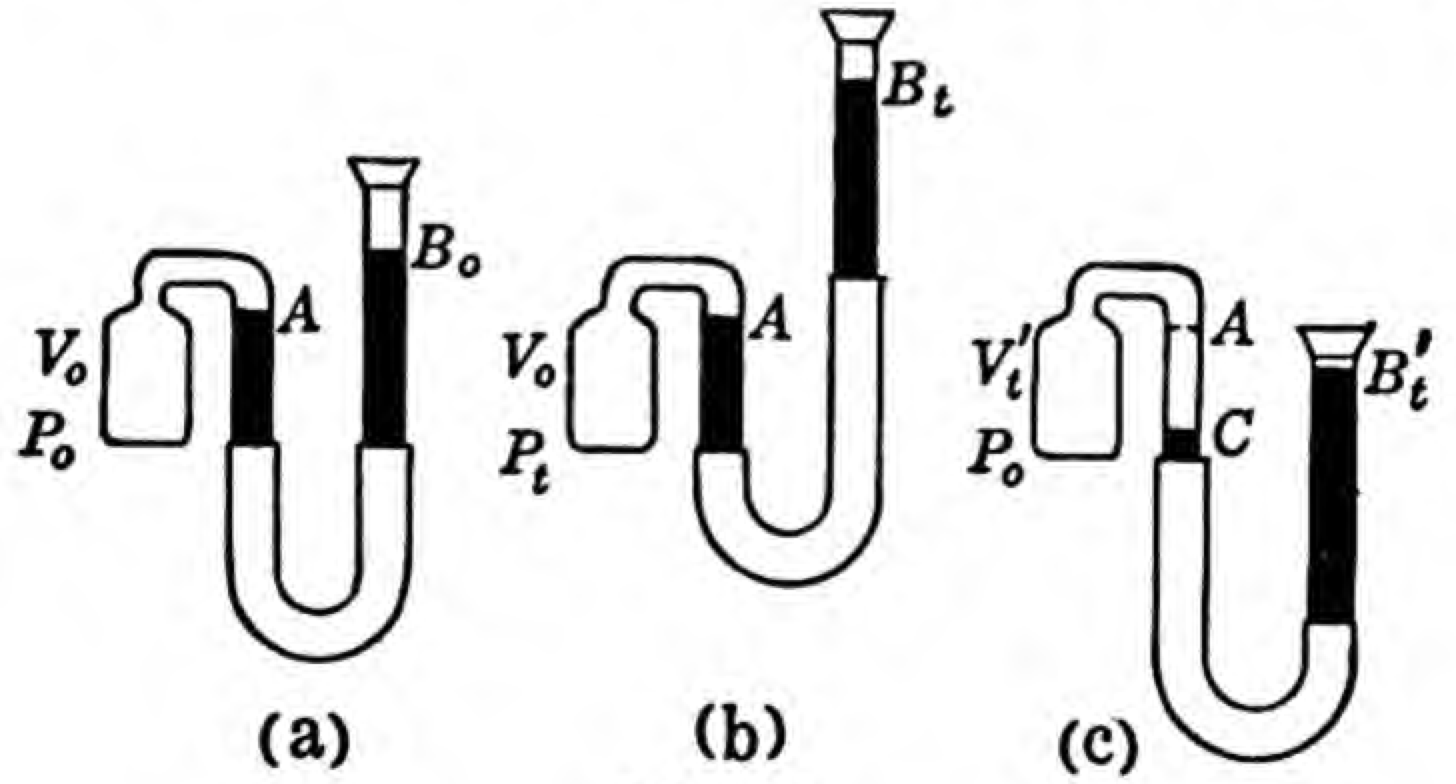


FIG. 60

The lower part of the tube is usually of rubber and enough mercury is employed to keep the levels on both sides well up into the glass upper part. When the bulb is immersed in a mixture of ice and water, let the mercury levels be at A and B_0 (Fig. 60a). The bulb is next transferred to a warmer bath (Fig. 60b). The right-hand tube is raised until the expanded gas is forced back into the

same volume with the mercury level again at A. The increase in pressure and thence the pressure p_t is determined from the difference in the levels B_t and B_0. From Eq. 119,

$$p_t v_0 = p_0 v_0 (1 + \alpha t) \tag{122}$$

for any value of t. Now leaving the bulb in the warmer bath, but using the same apparatus as a constant pressure thermometer, the right-hand tube is lowered until the difference in levels of C and B_t' (Fig. 60c) is the same as the original difference of A and B_0. The pressure of the gas is then the same as at the start. The value of $v_t - v_0$ is determined from the cross section of the tube and the distance AC. From Eq. 121,

$$p_0 v_{t'} = p_0 v_0 (1 + \beta t') \tag{123}$$

for any value of t'. The change represented by the diagrams (b) and (c) has been carried out at the same actual temperature, so Boyle's law may be applied: $p_t v_0 = p_0 v_t'$. Hence $\alpha t = \beta t.'$ But no limitation has been put on the temperature of the warm bath, so considering it to be at the boiling point where by definition t and t' are equal, we get $\alpha = \beta$. So, for a perfect gas, the pressure and volume coefficients are equal, and therefore the two temperature scales are the same.

Now at the given temperature we may expand the gas to any pressure p and volume v; hence, we may write

$$p_0 v_t = p_t v_0 = pv. \tag{124}$$

Therefore, by Eq. 122 or 123,

$$pv = p_0 v_0 (1 + \alpha t). \tag{125}$$

If v is kept constant and t is lowered, p decreases. Imagine that the temperature decreases to such a point that p decreases to zero (in the next article this will be shown to be reasonable). No lower temperature is conceivable on the gas scale and it is, therefore, called the absolute zero. When $p = 0$, Eq. 125 gives $t = (-1/\alpha)°$ C. Taking $\alpha = 0.003660$, we obtain $t = -1/\alpha = -273.2°$ C. Counting in Centigrade degrees from this absolute zero upward, the ice point is 273.2° Ab. and is called T_0 and any temperature $t°$ C. is $(273.2 + t)°$ Ab. Temperatures on the absolute gas scale are represented by T. Eq. 125 may be written

$$\begin{aligned} pv &= p_0 v_0 \alpha\left(\frac{1}{\alpha} + t\right) \\ &= p_0 v_0 \alpha (273.2 + t) = p_0 v_0 \alpha T = \frac{p_0 v_0}{T_0} T. \end{aligned}$$

Hence
$$\frac{pv}{T} = \frac{p_0v_0}{T_0}. \tag{126}$$

This is one common form of the equation of state of a gas and is called the *general gas law.* A gas which obeys this law is called a perfect gas.

The values of α and β for several gases are shown in the following table. When these values are changed so as to correct for the failure of the gases to obey Boyle's law the values of α and β become both equal to 0.003660.

	α	β
Air	0.003671	0.003674
Hydrogen	.003661	.003662
Nitrogen	.003673	.003672
Carbon Dioxide	.003728	.003712

The pressure p_0 for a given gas is directly proportional to M, the number of grams of gas enclosed. The proportionality constant depends on the kind of gas. Combining this constant with v_0/T_0 into a new constant r, Eq. 126 may be written

$$pv = MrT. \tag{127}$$

The value of r is different for each gas. This very common form of the gas law is not usable unless one has a table of values of r for all the different gases. A more universal form will next be obtained.

Consider several containers of equal volume filled with different gases to the same pressure when the temperatures are all equal, *i.e.* $v_1 = v_2 = v_3 = \cdots$; $p_1 = p_2 = p_3 = \cdots$; $T_1 = T_2 = T_3 = \cdots$. Avogadro's law (first stated as an hypothesis) states that under these conditions there are the same number of molecules (not atoms) in each container. Therefore, the mass of gas in each container is proportional to the mass of a molecule in each container, which, in turn, is proportional to the molecular weight. So we conclude that the ratio of the mass M of each gas to its molecular weight w is the same for all gases. We shall call this constant ratio N.

$$\frac{M_1}{w_1} = \frac{M_2}{w_2} = \frac{M_3}{w_3} = N.$$

The number of grams of a gas equal to the molecular weight of that gas is called one gram-molecule, or one mole. Thus one mole of hydrogen (H_2) is 2 gms.; of helium (He) is 4 gms.; of oxygen (O_2) is 32 gms.; etc. The ratio M/w is seen to give the number of

moles, N. Therefore *equal volumes of gases at the same temperatures and pressures contain equal numbers of moles of gas.*

Now rewrite Eq. 127 so as to have N appear in it:

$$pv = \frac{M}{w} wrT = N(wr)T.$$

When we take conditions such that the p's, v's, and T's are the same, the N's are then the same. It follows that the product wr is the same no matter what the gas. This product, $wr = R$, is called the *universal gas constant.* Then

$$pv = NRT, \tag{128}$$

where N is in moles. Sometimes the equation is written $pv = RT$, where R is still the universal gas constant, but necessarily v must be interpreted as v/N, or the volume occupied by one mole. The value of R is 8.314×10^7 ergs per mole per degree C.

Hydrogen is very nearly a perfect gas as far as following Boyle's law is concerned. On this account the scale of the constant volume hydrogen gas thermometer has been chosen by international agreement as the standard temperature scale. The corrections necessary to reduce the readings on this scale to those on the perfect gas scale are small, and therefore temperatures may be very accurately determined by its use. As will be proved in § 96, the corrected readings on the gas scale are identical with the absolute thermodynamic scale. This latter scale is independent of any material substance.

70. Work on the *p-v* Diagram. — When the piston (Fig. 61) compresses the gas in the cylinder of cross-section a, and moves a distance ds, work $dW = F \cos \theta \, ds$ is done. From the definition of pressure, $F = pa$. The change of volume dv is equal to $a\,ds$. Therefore $dW = pa\,ds = p\,dv$. On the (p, v) diagram, $p\,dv$ is the area of the narrow shaded strip of width dv and height up to the curve. The work done by the gas in expanding from volume v_1 to volume v_2 is

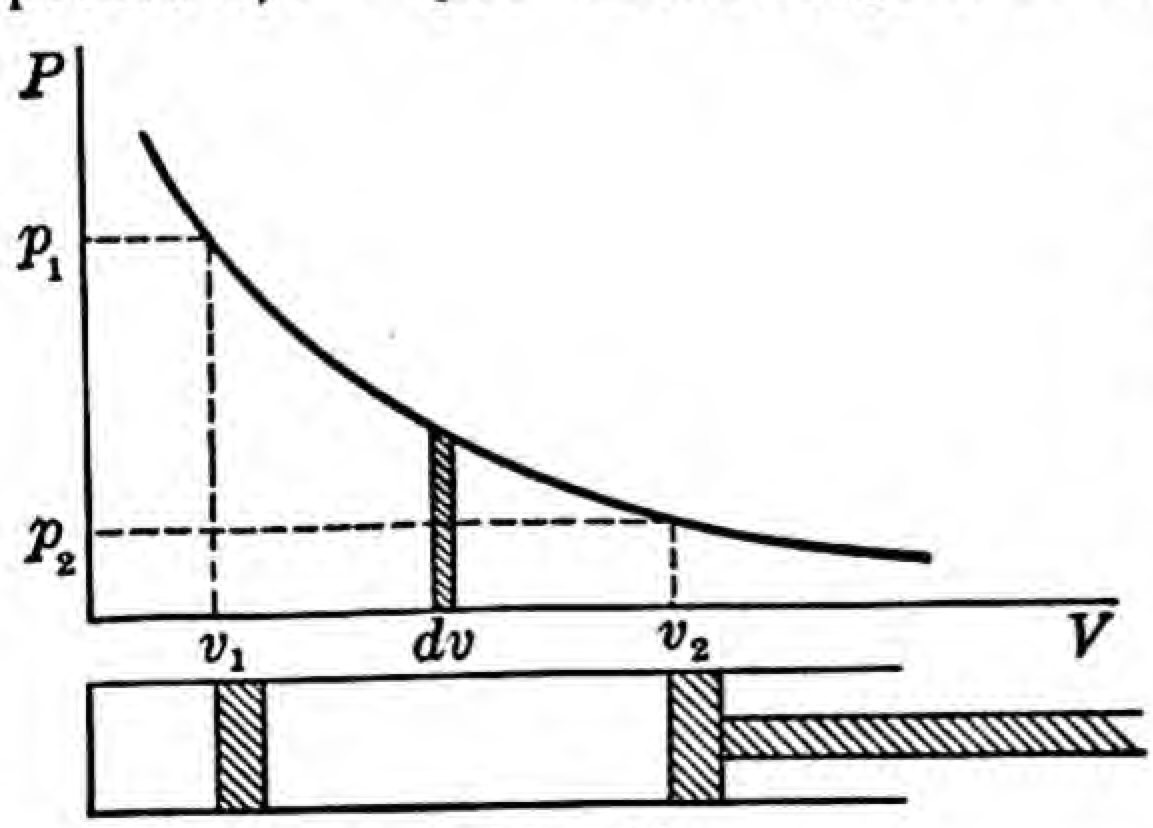


FIG. 61

$$W = \int_{v_1}^{v_2} p\,dv. \tag{129}$$

This expression may be integrated as soon as p is expressed as a function of the volume v.

If the expansion occurs isothermally (at constant temperature) at the temperature T_1, the expression in Eq. 129 may be integrated as follows: By Boyle's law, $pv = p_1v_1 = p_2v_2 = NRT_1$. Under standard pressure p_0 and at 0° C., $p_0v_0 = NRT_0$. Hence

$$
\begin{aligned}
W &= \int_{v_1}^{v_2} \frac{p_1v_1}{v}\,dv = p_1v_1 \log_e \frac{v_2}{v_1} \\
&= 2.3026\, p_1v_1 \log_{10} \frac{v_2}{v_1} \\
&= 2.3026\, \frac{p_0v_0T_1}{273.2} \log_{10} \frac{v_2}{v_1} \\
&= 2.3026\, NRT_1 \log_{10} \frac{v_2}{v_1}. \qquad (130)
\end{aligned}
$$

If we choose to compress the gas by doing work *on* the gas, we have to integrate from v_2 to v_1. Then W will turn out to be negative because every value of dv will be negative. We therefore agree to call positive the work done *by* the gas against outside forces and call negative the work done *on* the gas by an outside force.

QUESTIONS AND PROBLEMS

1. Taking the density of mercury at 0° C. as 13.595 gms. per cm^3. show that the standard atmosphere of pressure is 1,013,200 dynes per cm^2.

2. Describe the experiment necessary for determining where the scale divisions should be located in calibrating a constant volume gas thermometer so that it will read Centigrade temperatures. Describe the same for a constant pressure gas thermometer.

3. A barometer tube is filled with mercury to within one inch of the open end. It is then inverted, and set in a trough of mercury. After inversion, the air expands and occupies 12.00 inches of the tube and the mercury stands 27.00 inches above that in the trough. Find the true height of the barometer. *Ans.* 29.45 inches.

4. At the temperature of water boiling under a pressure of 760 mm. of mercury, the hydrogen in a hydrogen constant volume thermometer bulb is under a pressure of 730 mm. When the bulb is placed in a bath of melting metal the pressure is increased to 1432 mm. What is the Centigrade temperature of the melting point of the metal?

5. Find the work done when 10 grams of hydrogen at 0° C. and initially at two atmospheres pressure expands without temperature change to four times the initial volume. What change in pressure occurs? (The density of hydrogen under standard conditions is 0.000,089,880 gms. per cm^3.)

6. Compute the work necessary to compress a gas originally having a volume of 20 liters and a pressure of 76 cm. of mercury to 3 liters, the temperature remaining constant. Express the result in joules.

7. Calculate the volume of a gram-molecule of hydrogen under standard conditions.

8. How much external work is done when 10 liters of a gas at 20° C. are heated to 100° C., keeping the gas at constant pressure of 76 cm. of Hg.?

9. Change the value of the universal gas constant (given in § 69) into the number of B.T.U.'s per lb.-molecule per degree F.

The Kinetic Theory of Gases

71. The Assumptions of the Kinetic Theory. — Chemistry teaches that a finite amount of any substance is made up of an enormous number of very small discrete particles called molecules, which for the same substance are all alike in form, size, and mass. The kinetic theory of gases assumes further that, in the case of a gas of not too great density, the actual bulk (volume) of the molecules is very small as compared with the volume occupied by the gas; that when the gas is in "equilibrium," the molecules are distributed with approximate uniformity throughout the container; that they are moving about in all directions with different individual velocities; and that when they bump into the walls of the container, they rebound without any loss of energy (*i.e.* with perfect elasticity) and thereby exert a pressure against the walls. Although the individual velocities differ widely and the velocity of any one molecule changes from time to time, it is assumed that the average velocity for a sufficiently large number of molecules remains constant.

These are, briefly stated, the essentials of the kinetic theory of gases. Though it is properly called a theory, yet its agreement with all available experimental facts is such that it stands today unquestioned.

72. Pressure and Temperature in Terms of the Kinetic Theory. — In order to simplify the reasoning we will make two artificial assumptions known to be untrue, namely, that the molecules are smooth spheres and that the walls upon which they impinge are perfectly smooth so that a molecule striking normal to the plane of the surface is reflected back upon itself. We may justify these assumptions by stating that in the aggregate the effect is the same as though the molecules were smooth spheres no matter what their shape and, considering the aggregate of impacts, the

effect is the same as though the walls were mathematically smooth. The fact that the results deduced on this basis agree with experimental facts indicates that in the aggregate the effect is the same as though these conditions were actual.

Consider a gas enclosed in a cube the distance along whose edges is s. The molecules are traveling haphazardly in all directions. We may resolve the velocity of each molecule into its components along the x, y, and z axes. If the actual velocity of a certain molecule is V_1, the components are u_1, v_1, w_1 and

$$V_1^2 = u_1^2 + v_1^2 + w_1^2.$$

Consider now the component u_1. When a molecule of mass m strikes the cube face, it will be reflected with a velocity u_1, in the reverse direction, the change of momentum being $2\,mu_1$. The actual forces on the molecule and the wall are zero except when the molecule is in collision with the wall. While one given molecule is not colliding with the wall, other molecules are doing so, so that an average pressure, sensibly uniform, is exerted on the wall. So we are justified in considering that the change in momentum occurs gradually throughout the whole time t_1 between collisions. Let f_1 be the average force thus produced by the molecule of velocity u_1. Then $f_1t_1 = 2\,mu_1$. The molecule will travel across the cube and back again in the time $t_1 = 2\,s/u_1$ and hence the average force exerted on the face due to the single molecule is

$$f_1 = \frac{2\,mu_1}{t_1} = \frac{mu_1^2}{s}.$$

If we add up the effects on this face for all of the molecules, we have a total average force

$$F = \frac{m(u_1^2 + u_2^2 + \cdots)}{s}.$$

Since the average pressure p is the average force per unit area,

$$\frac{F}{s^2} = p = \frac{m(u_1^2 + u_2^2 + \cdots)}{s^3},$$

and we may write $$p = \frac{nm(u_1^2 + u_2^2 + \cdots)}{s^3n},$$

where n is the total number of molecules in the cube. Let us write

$$\frac{u_1^2 + u_2^2 + \cdots}{n} = \overline{u}^2,$$

where $\overline{u}^2$ is the average of the squares of the x-velocity components.

Therefore, $$p = \frac{nm\overline{u}^2}{s^3}. \tag{131}$$

Since $s^3 = v$ the volume, $pv = nm\overline{u}^2 = M\overline{u}^2$, where M is the total mass. Because the pressure is the same along all the axes, it follows that $\overline{u}^2 = \overline{v}^2 = \overline{w}^2$. Therefore $\overline{u}^2 = \overline{V}^2/3$, where $\overline{V}^2 = \overline{u}^2 + \overline{v}^2 + \overline{w}^2$ is the average square of the velocity of all the molecules. So

$$pv = \tfrac{1}{3}\, nm\overline{V}^2 = \tfrac{1}{3}\, M\overline{V}^2. \tag{132}$$

The student should write down the first few terms of the expression representing the kinetic energy of linear motion of each molecule along each axis and show that the sum of these reduces to $\frac{1}{2}\, M\overline{V}^2$.

In § 69 we defined a temperature scale such that $pv = NRT$. Combining this equation with Eq. 132 we obtain

$$\tfrac{3}{2}\, NRT = \tfrac{1}{2}\, M\overline{V}^2 = \text{Kinetic energy of translation of all of the molecules.}$$

Therefore from the simple assumptions as to the behavior of an ideal gas we have deduced that there should be a direct proportionality between the total kinetic energy of linear motion of a gas and the absolute temperature of a gas.

No attempt has been made to take into account the kinetic energy of rotation of the molecules, of vibrational energy between atoms inside the molecule, or the internal potential energy of the molecules. Analysis too complicated to be reproduced here shows that the same type of relation as derived above should hold for the total energy content of a gas.

73. Avogadro's Law and Dalton's Law of Partial Pressures. — Equal volumes of different gases at the same pressure and temperature contain the same numbers of molecules (§ 69).

We may write Eq. 132 as

$$pv = \frac{1}{3}\, nm\overline{V}^2 = \frac{2}{3}\, n\, \frac{m\overline{V}^2}{2}.$$

For one gas $$p_1v_1 = \frac{2}{3}\, n_1\, \frac{m_1\overline{V_1}^2}{2};$$

for a second gas $$p_2v_2 = \frac{2}{3}\, n_2\, \frac{m_2\overline{V_2}^2}{2}.$$

If the pressures of two gases are equal and their volumes are likewise equal, we see from the above equations that

$$n_1 \frac{m_1 \overline{V_1}^2}{2} = n_2 \frac{m_2 \overline{V_2}^2}{2}.$$

At any given instant of time a certain few molecules are at rest, while others have velocities of various magnitudes. The next instant, due to collisions, the velocities of the individual molecules are widely different. However, if no energy has been added to the gas, the total kinetic energy of all the molecules is not changed. Therefore if we divide the total kinetic energy of the gas by the number of molecules present, we get a quantity $m\overline{V}^2/2$ which we may call the average kinetic energy of a given molecule. Now it may seem plausible that in all gases at the same temperature the average kinetic energy of the molecules would be the same. Assuming this to be true, we then deduce that

$$n_1 = n_2.$$

Because there is much independent chemical and physical evidence of the correctness of Avogadro's law, it seems better to reverse the above logic and say that from Avogadro's law and from the reasoning of kinetic theory we deduce that the average energy of a molecule is the same for all gases if the temperatures of the gases are the same.

Dalton studied the pressure exerted by mixtures of gases. He found that if certain amounts of two different gases were separately put into a given container and the pressure of each measured, then when the two were put into the container together, the pressure was the sum of the two individual pressures. That portion of the total pressure in a container which is due to a certain gas (which would be the pressure in the container if that gas were alone present) is called the partial pressure of that gas. Dalton stated that in a mixture of gases, each gas exerts its own partial pressure independent of the presence of any other gases.

The student should see whether or not Dalton's law follows from the results of the previous work.

Variations from Dalton's law may be expected if the components of the mixture of gases or vapors react (more or less completely) chemically. In fact, we might reverse matters and say that the degree of failure of Dalton's law should show the degree of chemical or physical action between the components of the mixture.

74. Brownian Movement. — While it is physically impossible to see the individual molecules and study their motions, direct optical evidence of the existence of such motion is to be found in the so-called Brownian movements. If a drop of water containing any very finely divided material in suspension, such as, for example, carmine, is examined under a high-power microscope, the individual particles will be seen to possess a peculiar jiggling, dancing motion. This motion is considered to result from the molecular bombardment to which they are being subjected. The particles are huge compared with the molecules, but are caused to move in a zig-zag path under the innumerable impacts. A log floating in the water and struck intermittently by rifle bullets from all sides would be knocked about in a similar way.

75. Deviations from Boyle's Law — Van der Waals' Equation. — Careful experiments show that the general gas law is not exactly true over large ranges of pressure or temperature. Van der Waals was among the first to modify the simple assumptions of § 71. He assumed that if more careful study was made of the actual mechanism of the molecular motion, the corrected formula would still be of the same mathematical form, *i.e.* some idealized pressure multiplied by an idealized volume would be proportional to the absolute temperature. The molecules actually must occupy an appreciable space and it is only the space between them which undergoes change with pressure. The observed volume v therefore is greater than the variable or idealized volume by some constant amount, b. The idealized volume is then $v - b$. Further, the idealized theory takes no account of the attractions between the molecules. Consider a molecule close to a unit area anywhere in the gas. The force acting on that molecule due to the attraction by all the other molecules of the gas will vary directly as the number of the other molecules or, in other words, directly as the density of the gas. The number of molecules near the unit area which are thus acted upon also varies directly as the density of the gas. Therefore the attraction between all the molecules near the unit area and the other molecules varies as the square of the density of the gas. This attractive force per unit area causes a decrease in the gas pressure, and because of it the measured pressure will be lower than that computed on the basis of mechanical impacts alone. Therefore the idealized pressure will be higher than the actual pressure by an amount proportional to the square

of the density. To get the ideal pressure we must add to the observed pressure an amount $c\rho^2$ or a/v^2, where c and a are constants, ρ the density, and v the volume of the gas. The equation of state then becomes

$$\left(p + \frac{a}{v^2}\right)(v - b) = MrT. \tag{133}$$

This is known as Van der Waals' equation. Although this equation is in much closer agreement with experiment than the general gas law, it does not give complete agreement. Several other modifications of the perfect gas law have been made.*

If the product pv be plotted against p, we would get according to Boyle's law a straight line parallel to the p axis. According to Van der Waals' equation, we would get curves as indicated in Fig. 77. Curves of this sort are actually obtained experimentally.

PROBLEM

1. Compute the value of $\sqrt{\overline{V^2}}$ for a molecule of hydrogen under standard conditions. Also for oxygen.

Heat as Energy

76. Quantities of Heat and the First Law of Thermodynamics. — For many years in the early development of the science of Physics, the true nature of heat was not understood. Before Joule had proved that heat is only a particular manifestation of energy, heat was known only through the temperature changes which it produced. Therefore the following units were selected.

Considering the average laboratory temperature to be about 15° C., the *calorie* was defined as the amount of heat required to change the temperature of 1 gm. of water from $14\frac{1}{2}°$ to $15\frac{1}{2}°$ C., which change is symmetrical about the 15° point. This unit is now called the *Fifteen Degree Calorie*, and has abbreviations cal_{15}.

* Clausius assumed that the attractive force between the molecules varies with temperature and depends in a manner on the density. The Clausius equation is

$$p = \frac{RT}{v - b} - \frac{a}{T(v + c)v}.$$

Dieterici has suggested the following equation of state:

$$p(v - b) = RT\, e^{-\frac{a}{vT^n}}.$$

All of these various equations fail to fit exactly the actual experimental values. For certain uses one equation will be in close agreement with experiment; for other uses another equation is more desirable.

and C_{15}. Similarly a *Twenty Degree Calorie* has been proposed and used. The *Mean Calorie* is 1/100th part of the heat necessary to raise one gram of water from 0° C. to 100° C. It differs from the other calories slightly.

The *B.T.U.* (*British Thermal Unit*) is 1/180th part of the quantity of heat necessary to raise one pound of water from 32° to 212° F. It is approximately equal to 252 calories.

The *heat capacity* of a body is the quantity of heat required to raise the temperature of the body one degree. The *water equivalent* of a body is the amount of water which has the same heat capacity as the body.

The *specific heat* of a substance was originally intended to be the ratio of the quantity of heat necessary to raise the temperature of one gram of the substance one degree Centigrade to the quantity of heat required to raise one gram of water from $14\frac{1}{2}$° C. to $15\frac{1}{2}$° C. It was thus a pure number without units. Now it is usually defined to be the number of 15° calories necessary to raise one gram of a substance one degree Centigrade.

It follows from this definition that when a substance changes temperature (without a change in state) the quantity of heat necessary to produce the change is computed by the product of the mass m, the specific heat c, and the change in temperature Δt of the body.

We may speak of the specific heat of water itself. Thus the specific heat of water at 70° C. is the number of 15° calories of heat necessary to raise 1 gm. of water from $69\frac{1}{2}$° C. to $70\frac{1}{2}$° C.

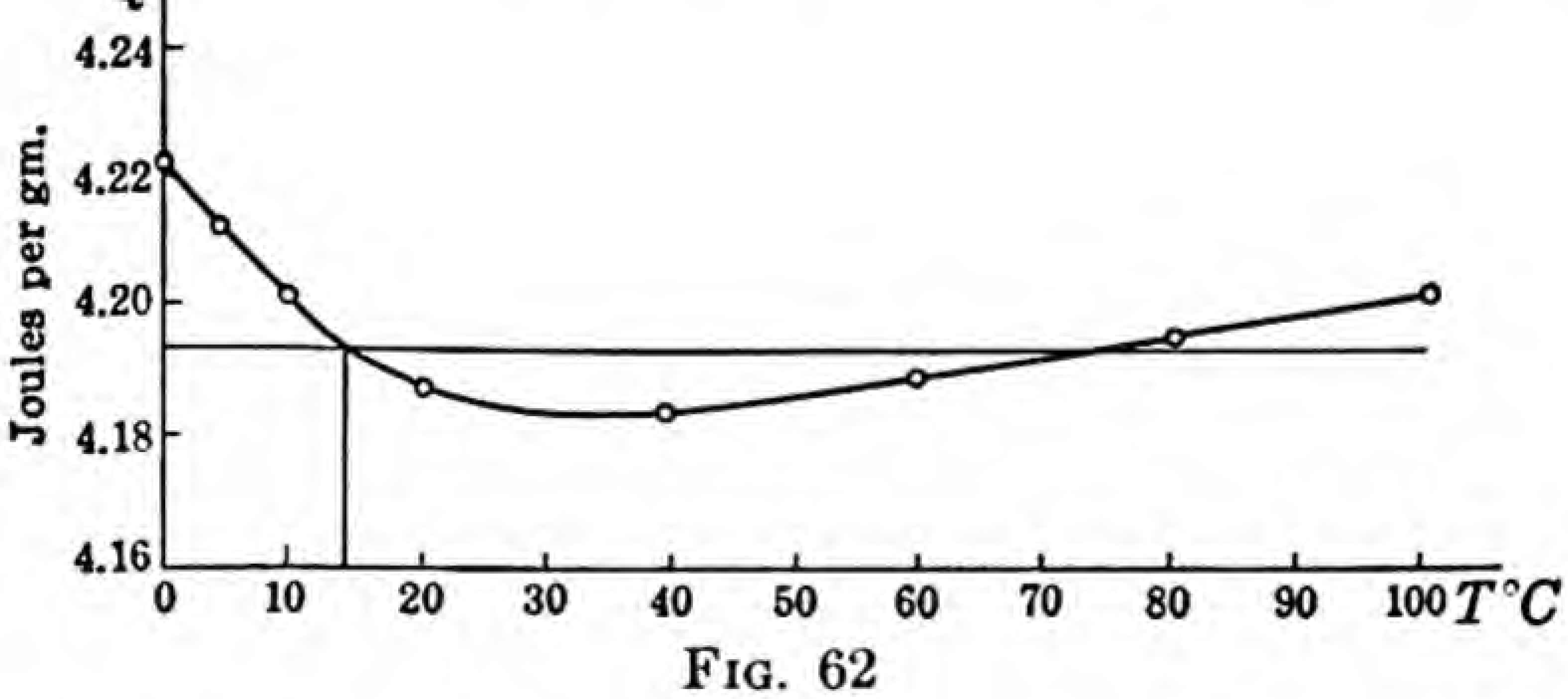


Fig. 62

As shown in Fig. 62, the specific heat of water near 0° C. is more than one calorie per gram per C°. It passes through the value of one calorie per gm. per C° at 15° C. and reaches a minimum at about 30° C. It does not reach quite as high a value at 100° as it had at 0° C.

Heat was first thought to be a fluid, called caloric. It was thought that a definite amount of it was in every body and that it could flow from one body to another body placed in contact with it. A body of greater specific heat than another was supposed to have a greater affinity for caloric than the other. Count Rumford, in 1798, published observations showing that the hypothesis according to which a body possesses a certain definite amount of caloric was not valid since by drilling in metal with a blunt drill, where almost no material was abraded, an almost unlimited amount of heat could be developed. Then Joule in 1842 constructed a calorimeter in such a fashion that he could make careful measurements of the rise in temperature and accurate measurements of the amount of work done in friction between parts of the calorimeter. He found that for every 41.9 million ergs of work done, there was one calorie of heat produced in the calorimeter. It was then obvious that the intangible thing called heat, which is manifest only by changes in temperature of a body and whose transfer from one body to another is inferred by the changes in temperature of two bodies, was nothing more than energy.

As soon as Joule's discovery was made, the principle of *The Conservation of Energy* was stated as applied to heat. This principle is called *The First Law of Thermodynamics*, and may be stated: whenever heat is converted into work a definite amount of work is always produced by a given quantity of heat and conversely. The latest experiments give 4.185 joules of work = 1 cal_{15}. The letter J is used to stand for 4.185 joules/cal_{15}. or 777.8 ft. lbs./B.T.U.

Any given body possesses a certain amount of potential energy due to the position of its molecules with respect to each other. The body will also have some amount of kinetic energy of random vibration of its molecules. The total amount of these two types of internal energy is called the *Intrinsic Energy* of the body and is designated as U. The whole quantity of energy U is difficult to determine and is seldom desired, but the change in U, dU, is the thing that usually is of interest and is easily measured by the change in temperature or by change of state. Thus when a quantity of heat is added to a body, either the kinetic energy or potential energy or both may be increased (the substance may expand and thus have its molecules separated further and the temperature may rise) and if the body expands against the atmos-

phere or other external forces, work will be done. This is usually stated

Heat added = Change in Internal Energy + External work done in expansion,

or

$$dQ = dU + dW. \tag{134}$$

When referring to just one gram of the substance, we use small letters thus:

$$dq = du + dw.$$

This equation is a direct application of the First Law of Thermodynamics. In Eq. 134 it is seen that dW is positive when the work is done *by* the body. Then work done *on* the body from outside forces will be negative. That is in keeping with the convention in § 70.

The units of every term in Eq. 134 must be the same. If dW and dU are in work units and dQ in calories, then we must write $JdQ = dU + dW$. Because of the great accuracy with which electrical quantities can be measured, nearly all heat experiments nowadays are performed electrically where the energy is directly computed in joules, and therefore it is now becoming general to express dQ in joules or K.W.H. In this text dQ and Q will be expressed in work units. The new International Critical Tables express all of the data on specific heats and other heat quantities in joules. By this procedure it will not be necessary to revise tables of accurately measured heat data every time someone redetermines the value of J with higher precision. Thus the latent heat of vaporization of water at 100° C. and standard pressure is given in the International Critical Tables as 2258 joules per gram. This is judged to be accurate to 0.1 per cent. If the older procedure were followed, this datum would be reduced to

$$\frac{2258 \text{ joules}}{4.185 \text{ joules/cal.}} = 539.5 \text{ cal.},$$

and if later experiments should prove that a more accurate value of J differed from 4.185 joules/cal. by more than 0.1 per cent, then the value 539.5 cal. would be wrong although the original measurements in joules were quite independent of the value of J.

PROBLEMS

1. Given that 1 lb. = 453.592 gms., prove that one B.T.U. is nearly equal to 252 calories.

2. Prove that 1 B.T.U. is equivalent to 777.8 ft. lbs.

Change of State

77. Vaporization. — The phenomena of change of state, solid to liquid or liquid to vapor, may be readily explained in terms of the kinetic theory. In the liquid state the molecules are supposed to be so close together that their freedom of motion is restricted by molecular attractions. The molecules within the mass cannot break loose from the attractions of those surrounding them. The molecules have a certain average velocity determined by the temperature of the body as a whole, but each molecule has a variable velocity determined by collisions with its neighbors. When molecules near the surface attain sufficient velocity directed toward the surface, they may break away from the liquid, forming a vapor in the region above the liquid. These molecules constituting the vapor travel haphazardly in all directions and as a result a certain proportion must be continually re-entering the liquid. Thus we have molecules going both ways through the surface. At any particular temperature the escape of molecules through the surface will go on until a vapor pressure is established such that the number returning just balances those escaping. If the temperature of the liquid is raised, the average molecular velocity within the liquid increases and more molecules escape through the surface. This continues until the vapor pressure above the liquid again balances that within. A vapor in contact with a plane surface of its liquid and in equilibrium with it is called *a saturated vapor.* In order to have this condition of equilibrium, we must consider that there exists in the liquid a vapor pressure equal to the pressure of the saturated vapor in contact with it. This condition will obviously be modified if the surface of separation is not flat, — because of the surface tension of the liquid. The pressure of a saturated vapor depends upon the temperature but not upon the volume. If the temperature of a saturated vapor is kept constant, any attempt to compress the vapor results in sufficient condensation to maintain the pressure constant.

When a bubble of vapor is formed in a liquid, it is the bombardment pressure, due to the molecular impacts from within, which holds back the liquid walls. Molecular attraction in the form of surface tension tends to make the bubble collapse. The actual resultant pressure within the bubble, the saturated vapor pressure, must be equal to atmospheric pressure plus the hydrostatic pres-

sure plus the pressure due to surface tension, if the bubble is to persist.

When the temperature is such that the vapor pressure equals the external pressure, we are at the "boiling point" of the liquid. Any input of heat then produces change of state without rise in temperature until the change is entirely effected. It is found that the boiling point (temperature) is a function of the pressure. The pressure above a liquid is usually that due to the atmosphere; hence, usually we mean by "boiling point" the temperature at which the vapor pressure equals the atmospheric pressure.

The pressure of saturated water vapor as a function of the temperature is indicated by line OA (Fig. 63). This curve may also be interpreted as showing the boiling points for various external pressures and shows the fact that the boiling point may actually be lowered to nearly 0° C. by simply reducing the external pressure.

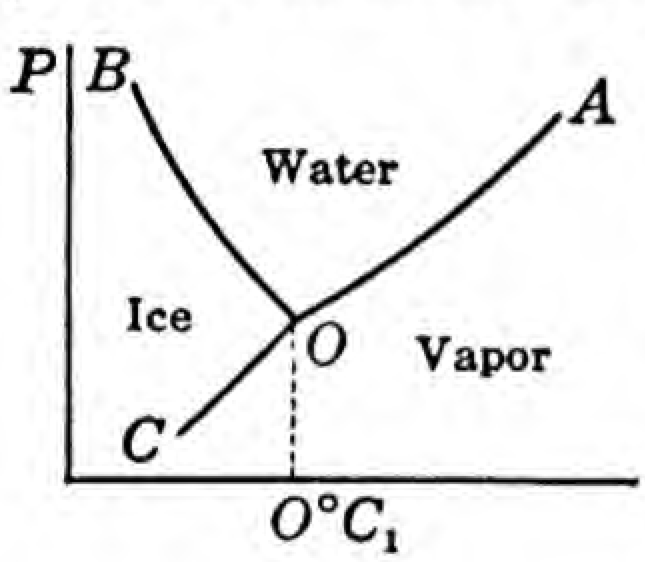


FIG. 63

In the liquid state the molecules are close enough together to be influenced by their mutual molecular attractions. Therefore, to change a substance from the liquid to the vapor state without any change in temperature, work must be done against these intermolecular forces. No change in the kinetic energy of the molecules occurs because there is no change in the temperature. Work must be done against the external pressure (saturated vapor pressure) because a very great increase in volume occurs. The heat required to change one gram of a substance from a liquid to vapor at the same temperature is called the *heat of vaporization*. This heat (energy), as just explained, is used partly in the performance of internal and partly in performance of external work. Eq. 134, when applied to this case, becomes

$$q_v = \Delta u_p + \int p\,dv, \qquad (135)$$

where q_v is the heat of vaporization, Δu_p the increase of the potential energy, and $\int p\,dv$ the external work of expansion for each gram of the substance.

As was seen in Fig. 63, the pressure at which boiling occurs decreases as the temperature decreases. At lower temperatures

the molecules are closer together, the molecular forces are greater, and hence Δu_p is greater. In order to know whether $\int p\,dv$ increases or decreases, actual data must be employed because as the temperature decreases the value of p decreases but the value of Δv, the volume occupied by one gram of the vapor at the pressure p, increases greatly. In the data given in Problem 9 below, it will be seen that for saturated steam $p\Delta v$ decreases as the temperature drops. Therefore, since q_v for saturated steam increases as the temperature drops, we conclude that Δu must increase at a greater rate than $p\Delta v$ decreases. Examination of the answers in Problem 9 proves this to be so.

It follows from the definition of the latent heat of vaporization that the quantity of heat necessary to vaporize m grams of a substance, *without any change in temperature*, is equal to the product of the mass m and the latent heat q_v.

78. Fusion. — In the solid state we have a still further restriction of the motions of the molecules. The crystalline structure usually characteristic of this state indicates a definite arrangement of molecules in positions of equilibrium under very large molecular forces. The molecules are still in motion, but this motion is no doubt largely vibration about equilibrium positions.

When heat is added to a solid the temperature rises until the melting point is reached, after which the further addition of heat causes change of state (fusion) without rise in temperature until the change is entirely effected.

The quantity of heat necessary to change a unit mass of a substance from the solid to the liquid state without change of temperature is called the latent heat of fusion, q_f, of the substance. It is very important, when computing the quantity of heat necessary to produce a given change of temperature, to inquire as to whether or not a change in state occurs in the region of change. Thus, let us compute the heat necessary to change 10 gms. of copper and 10 gms. of water each from $-10°$ C. to $105°$ C. For the copper there is no change in state and so the answer is $10 \times 0.09 \times [105 - (-10)] = 103.5$ calories. For the water there are two changes of state and we must compute the heat in 5 portions thus (see Problem 6 for data):

$$10 \times 0.50 \times [0 - (-10)] + 10 \times 80 + 10 \times 1 \times (100 - 0) + 10 \times 540 + 10 \times 0.48 \times (105 - 100) = 7274 \text{ calories.}$$

For the case of fusion, Eq. 134 becomes

$$q_f = \Delta u_p + \int p\,dv, \tag{136}$$

where Δu_p is the internal energy per gram required to tear apart the molecules of the solid (which is all an increase in potential energy and includes no increase in kinetic energy since the temperature has not changed); $\int p\,dv$ the external work due to the change in volume of one gram of the substance. $\int p\,dv$ may be either positive or negative, depending upon whether fusion entails increase or decrease in volume. It is very small compared with Δu.

Even in the case of a solid there still exists a definite though small vapor pressure, which is due to molecular motion. There is probably an interchange of molecules from group to group, one molecule replacing another without disturbing the grouping, and hence molecules escape through the surfaces and re-enter as in the case of a liquid. There exists a definite equilibrium pressure between the saturated vapor and the solid for each temperature. The saturated vapor pressure curve for solid and vapor in the case of ice is shown in Fig. 63, line OC.

The line OB (Fig. 63) indicates pressures at which ice and water may exist in equilibrium for various temperatures.

The three curves meet at a common point O, called *the triple point.* At this point the three phases, ice, water, and vapor, are all in equilibrium with each other. The temperature is a very small fraction of a degree above 0° C. The pressure is about 4 mm. of mercury. At any temperature and pressure represented by points to the left of BOC, the material will be ice. For conditions represented by points above BOA only water exists. Below COA only vapor can exist.

PROBLEMS

1. Solved Problem: The following substances are placed in a calorimeter. Determine the resulting temperature of the mixture:

Substance	Mass	Temp.	Specific Heat in cal./gm. deg. C.
Water	200 gms.	20° C.	1.
Copper	30 gms.	90° C.	.09
Ice	60 gms.	− 13° C.	.5

There are two common ways of working problems of heat mixtures. In one method an algebraic equation is written down and solved. Before an equation may be written down for the above problem it is necessary to know whether the temperature of the mixture is below, at, or above 0° C. Suppose that the final temperature is $t°$ above zero. Then we equate the heat given out by the copper and the water in cooling down to $t°$ to the heat received by the ice in warming up to $t°$.

$$200(20 - t) + 30 \times 0.09(90 - t) = 60 \times 0.50 \times 13 + 60 \times 80 + 60\,t$$

From which $t = -3.6°$ C. Since t was assumed to be above zero, our original supposition was wrong. Similarly if t were assumed to be below zero, the answer would be contradictory. The only other alternative is that $t = 0$ and that some ice and water exist in equilibrium. Let m grams of ice be melted. Then

$$200 \times 20 + 30 \times 0.09 \times 90 = 60 \times 0.50 \times 13 + m \times 80$$
$$m = 48.2 \text{ gms.}$$

Hence the final state consists of 30 gms. of copper, 248.2 gms. of water, and 11.8 gms. of ice, all at zero degrees Centigrade.

In order to avoid this cut-and-try method we proceed as follows. We compute the amount of heat which must be removed from the whole system in order to lower it to the lowest temperature of any part of the system, $-13°$ C. in the above case. Thus $200 \times 20 + 200 \times 80 + 200 \times 0.50 \times 13 + 30 \times 0.09(90 + 13) = 21{,}578$ calories. Now let this heat be added piecemeal to the system. The heat necessary to raise the system from $-13°$ C. to 0° C. is $260 \times 0.5 \times 13 + 30 \times 0.09 \times 13 = 1725$ calories. To melt all of the ice requires $260 \times 80 = 20{,}800$ calories. But only $21{,}578 - 1725$ or 19,853 calories are available. Hence only $19{,}853 \div 80$ or 248.2 gms. of the ice are melted, leaving 11.8 gms. unmelted.

Whenever the region of final temperature is obvious from inspection of the problem, the first method is the simpler, otherwise the latter method should be used.

2. If 100 grams of lead cool from 340° C. to 327° C. in 2 minutes and then the temperature remains steady for 25.8 minutes while the mass is solidifying, find the heat of fusion of lead, assuming that heat is lost at a uniform rate and that the specific heat of lead at 330° C. is 0.028 cal. per gram per degree C.

3. A 300 horsepower engine is given a brake test. The brakes are water cooled. At what rate must water at 80° F. flow through the brakes if the water must not rise above 180° F.?

4. A pyrex tube $1\frac{1}{2}$ meters long and containing 2 liters of water contains 1000 gms. of fine lead shot. The tube is turned end over end, letting the shot fall the full length of the tube. The weight of the tube is 1000 gms. and its specific heat is 0.20 cal./gm. How many times must this be repeated before the water rises 1° C.? Does the buoyant effect of the water upon the shot affect the answer?

5. What horsepower is required to raise the temperature of 100 Kg. of water at 20° C. to the boiling point in 30 minutes?

6. 12.5 Kg. of steam at 150° C. are mixed with 200 gms. of ice at − 30° C. Describe the state of affairs when equilibrium has been reached. Solve the problem also for the case of only 2 gms. of steam; 15 gms.; 45 gms.; and 90 gms.

Average specific heat of steam	0.48
Average specific heat of ice	0.50
Latent heat of fusion of ice	80 cal./gm.
Latent heat of vaporization of water	542 cal./gm.

7. How much energy is required to raise 1 Kg. of mercury from − 80° C. to vapor at 357° C.?

Specific heat of solid mercury	0.032
Specific heat of liquid mercury	0.033
Melting point	− 38.8° C.
Heat of fusion	2.82 cal./gm.
Boiling point	357° C.
Heat of vaporization	68 cal./gm.

8. 1 kilogram of mercury at − 80° C. (temperature of carbon dioxide snow) is placed in 500 gms. of water at 10° C. What is the resultant state?

9. The International Critical Tables give the following values for the latent heat of vaporization of water at several temperatures. Compute the amount of external work of evaporation at each temperature and then study the changes in the external and internal work which occur.

Temp. Centigrade	Vapor Pressure		q, in Joules/gm.	Specific Volume of Steam in cm³./gm.	Answers in Joules	
	mm. of Hg.	gms./cm².			$\int p\,dv$	Δu
0	4.58	6.23	2494	206,300		
50	92.3	125.5	2379	12,020		
100	760.	1033.	2258	1,671	169.2	2089
150	3569.	4852.	2107	392		
180	7514.	10216.	2003	194		

10. One gram of ice at 0° C. and one atmosphere pressure occupies 0.0907 cm³. more than one gram of water at 0° C. Compute $\int p\,dv$ for the change. *Ans.* − 0.0022 cal. (negligible compared to 80 cal.).

11. (*a*) A jar containing only water and its vapor is sealed up and put into a kettle of water which is brought to the boiling point. What will the pressure within it become?

(*b*) If the above jar is half full of water and half full of air at pressure of 76 cm. and sealed at 20° C. and then put into the kettle of water which is brought to the boiling point, what will the pressure within become, neglecting the expansion of the water and the glass?

12. Two liquids A and B are introduced into two barometer tubes, the temperature of each being the same. It is noticed (1) that in both cases a little of the liquid does not evaporate, (2) that the mercury in the tube containing A is more depressed than that in the tube containing B. Which liquid would you expect to have the higher boiling point? Why?

79. The *P-V-T* Diagrams and the Critical Temperature. — Fig. 64 shows typical isothermals for the liquid and vapor phases of a substance. At high temperatures the curves approach the equilateral hyperbola curves represented by pv = constant, such as Curve 1. As the temperature is lowered, the curves become distorted as in Curve 2. Finally a temperature may be reached which gives Curve 3. At any lower temperature, as the pressure is increased, a point A, Curve 4, is reached where the vapor begins to condense. The pressure cannot be increased because condensation decreases the number of molecules of the vapor so that the pressure remains constant. This condition continues until point B is reached, when all the vapor is condensed into liquid. Then since liquids are generally not very compressible, the curve rises sharply. The condition represented at C is called the *critical point.* The values at C are called the *critical pressure, critical volume,* and *critical temperature.* At no temperature above the critical temperature may a gas be liquefied under any pressure howsoever large. Below that point there occurs on all isothermals a region where a change of volume at constant pressure indicates condensation.

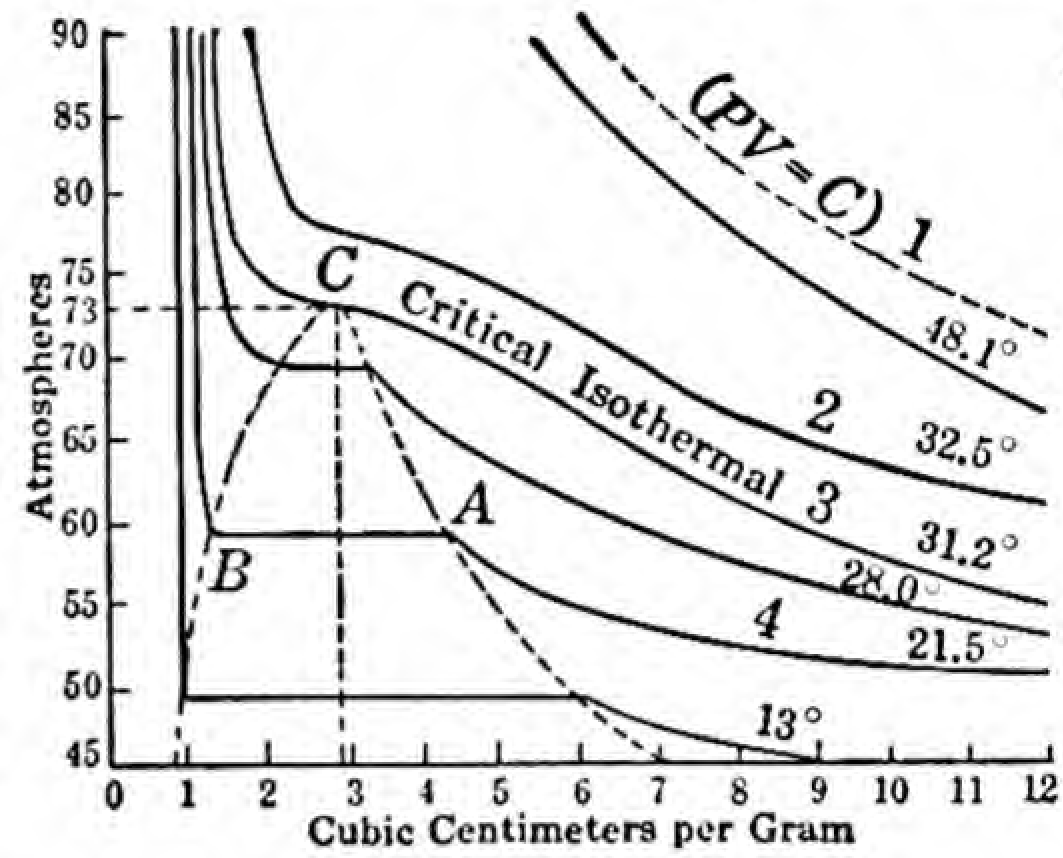


FIG. 64

The specific volume of a substance is the volume occupied by one gram of the substance. Fig. 65 shows the relation between the specific volume and the temperature for a saturated vapor and its liquid. The two curves meet at the critical point C. All distinction between vapor and liquid there disappears because both have the same density.

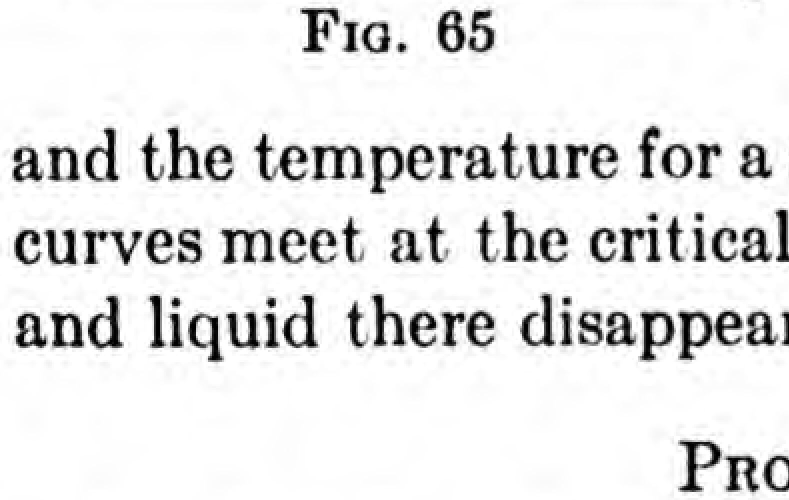


FIG. 65

PROPERTIES OF GASES

80. The Specific Heats of a Gas. — If a gas is allowed to expand while heat is being added, the temperature will not rise as rapidly as if the volume were kept constant because some of the energy added does the work of expansion. If the expansion is sufficiently

large, no change in temperature would occur and for still greater expansion a decrease in temperature would result. If the gas were compressed, less heat would be necessary to raise its temperature. Heat might actually be allowed to leave the gas and still have the temperature rise if the compression were sufficient. So the quantity of heat necessary to be added to a gram of a gas to raise its temperature one degree is positive, zero, or negative according to the expansion. Two of these various values are of special importance. The *specific heat at constant volume*, c_v, is the heat energy necessary to cause the temperature of one gram of the substance to rise 1° C., the volume remaining constant. The *specific heat at constant pressure*, c_p, is the heat energy necessary to cause the temperature of one gram of the substance to rise 1° C., the pressure remaining constant.

Let heat be added to M grams of gas in a closed container. The heat $Mc_v\Delta T$ is required to change the temperature of the gas by an amount ΔT. Since there is no work of expansion against either internal attraction or external pressure, all the energy must appear as increased kinetic energy of random motion of the molecules.

Now let heat be added to the same gas until the same change in temperature has occurred, but during that time let the volume increase so that the pressure does not change. The same amount of energy, $Mc_v\Delta T$, is required to give the same kinetic energy to the molecules regardless of their greater distance apart, but in addition there is required the energy $\int p\,dv$ to do the work of expansion against the constant external pressure and, if the gas is not a perfect gas, an amount of energy ΔU_p sufficient to separate further the molecules against their intermolecular forces. This latter energy is stored as molecular potential energy. All of these three amounts of energy would be included in the expression $Mc_p\Delta T$. We may then write

$$\Delta Q = Mc_p\Delta T = Mc_v\Delta T + \int p\,dv + \Delta U_p,$$

or

$$M(c_p - c_v)\Delta T = \int p\,dv + \Delta U_p. \tag{137}$$

Since the expansion is allowed to occur at constant pressure, then $\int p\,dv$ may be written $p\Delta v$.

For a perfect gas there is no molecular attraction, so ΔU_p would be zero. Then

$$M(c_p - c_v)\Delta T = \int p\,dv.$$

81. General Case of Heat Added to a Real Gas. — Let heat dQ be added to M grams of a real gas (possessing molecular attraction and hence internal potential energy) and let the pressure change in any prescribed way. If dT is the change in temperature of the gas, then we know from the previous section that the amount of internal kinetic energy required to produce that change in temperature is $Mc_v\,dT$. The rest of the energy acquired from the heat transfer must appear as external work and internal potential energy. So we know that

$$dQ = (Mc_v\,dT + dU_p) + p\,dv,$$

or

$$\Delta Q = (Mc_v\Delta T + \Delta U_p) + \int p\,dv. \tag{138}$$

The terms in the parentheses make up the total change in the internal or intrinsic energy.

PROBLEMS

1. Show that for a perfect gas

$$c_p - c_v = r, \tag{139}$$

where r is the gas constant for one gram of the gas in the equation $pv = MrT$. If $c_p - c_v$ should be expressed in calories, then $c_p - c_v = r/J$.

2. The curve in Fig. 66 shows the way water expands on heating. Which might be expected to be the greater between 0° C. and 4° C., c_p or c_v? Which the greater between 4° C. and 10° C.?

3. Discuss and write the energy equation for the case where the gas expands so much that the added heat produces no increase in temperature.

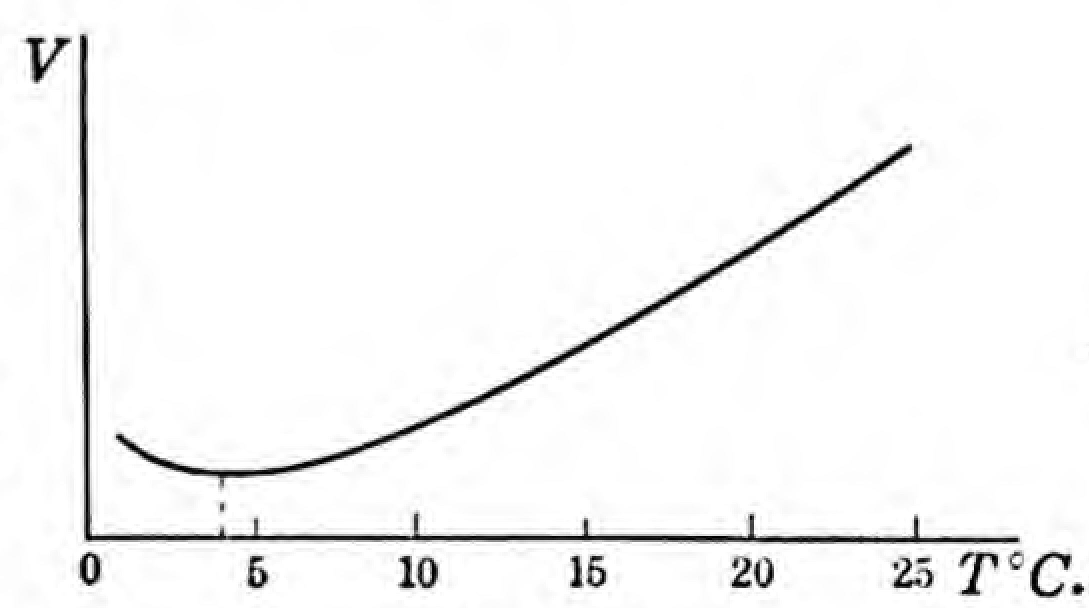


Fig. 66

82. Isothermal Changes. — If the pressure on the gas in a cylinder is changed, say by changing the force F (Fig. 67), the volume will change accordingly. If the change is made very slowly by infinitesimal steps, the flow of heat into or out of the cylinder

(assuming conducting walls) will keep the gas at the temperature of its surroundings. Such a change is an *isothermal change.*

Let the gas have the pressure p_2 and volume v_2, represented by the point c. If the pressure is increased very slowly to p_1, the values of p and the corresponding values of v map out the points on the curve between c and d. If the pressure is quickly increased from p_2 to p_1, the volume will be larger than v_1 but will slowly reach that value as the gas cools down to the original temperature. The isothermal curve AB is the locus of all the equilibrium states of p and v for a given temperature. For a perfect gas the equation of an isothermal curve is $pv = C$, C being a constant. If the whole process above is repeated at a higher temperature, the pressure at a given volume will be higher than in the first case and $pv = C$ still, but the value of C will be larger and the new isothermal curve will lie above the curve AB. From the general gas law we know that the value of C increases directly with the absolute temperature and with the mass of gas used.

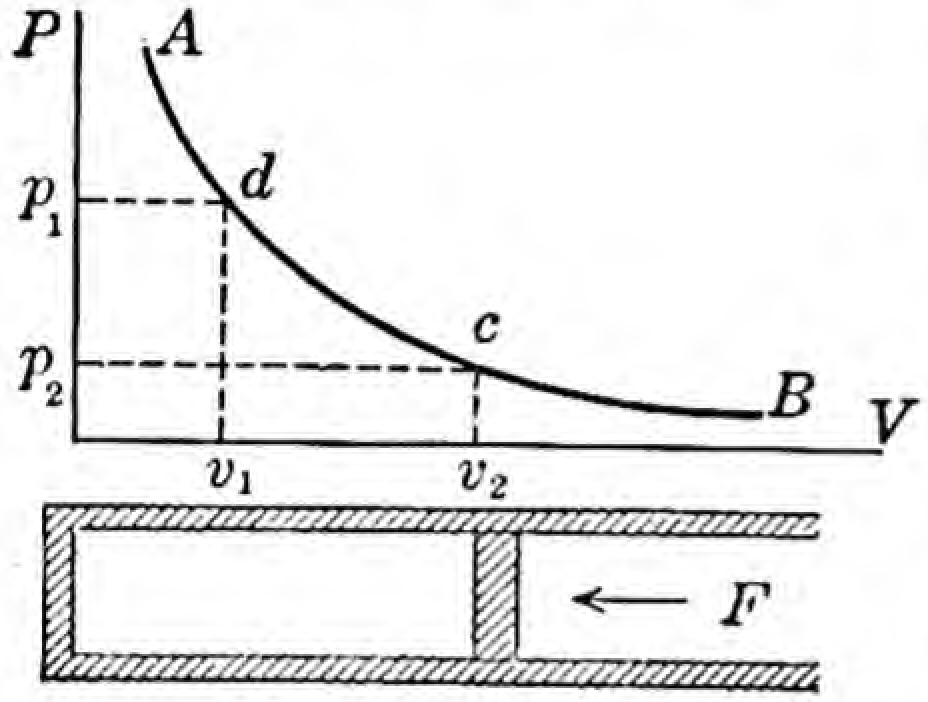


FIG. 67

83. Adiabatic Changes. — If the cylinder of gas shown above is insulated so that no heat can flow in or out, work done on the gas in compressing it must appear as an increase in the internal energy of the gas. Therefore an increase in pressure will produce a rise in temperature, and a decrease in pressure will then produce a drop in temperature. Such a change in which no heat is allowed to flow in or out is called an *adiabatic change.* If the coordinates of point c (Fig. 68) represent the initial conditions of the gas, an increase in pressure will cause the gas eventually to settle into equilibrium at some point d above the isothermal AB, since the temperature has risen. Thus the adiabatic curve is steeper than the isothermal. The equation for an adiabatic change for a perfect gas may be obtained readily from the expression for the First Law of

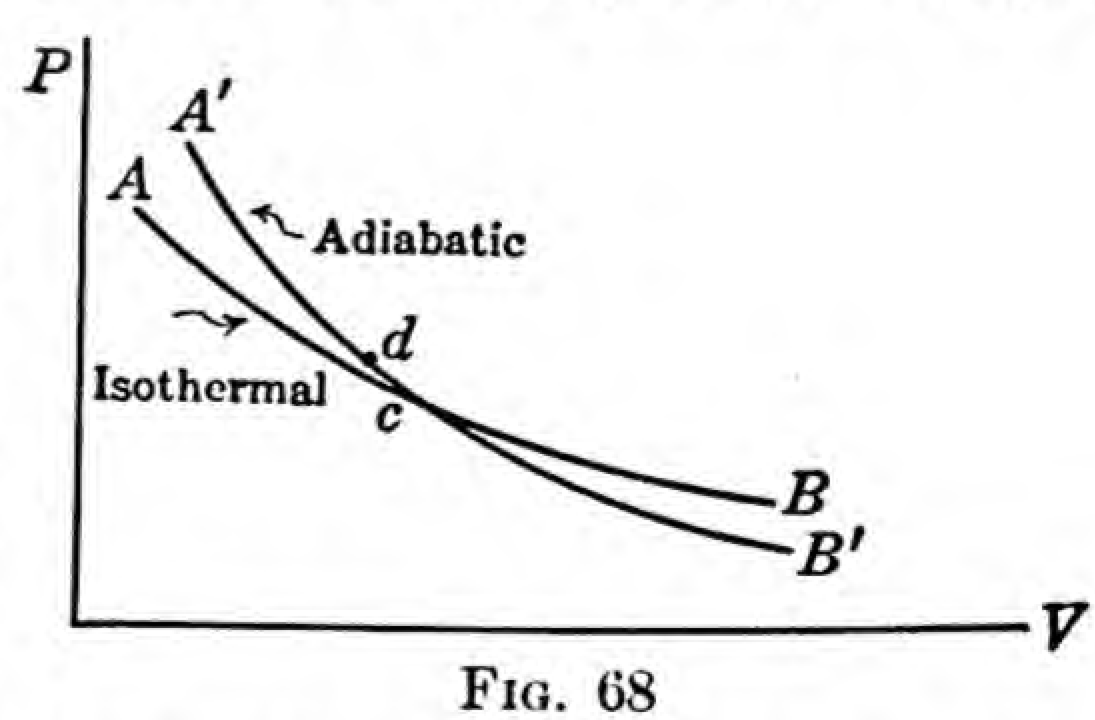


FIG. 68

Thermo-dynamics (Eq. 134). In this case $dQ = 0$, the internal potential energy is zero, and the change in internal kinetic energy is $Mc_v\,dT$. Then

$$0 = Mc_v\,dT + p\,dv. \tag{140}$$

Differentiating Eq. 127 and using Eq. 139, we obtain

$$p\,dv + v\,dp = Mr\,dT = M(c_p - c_v)\,dT.$$

Substituting the value of dT from this equation in Eq. 140, cross multiplying, canceling, and arranging,

$$\frac{dp}{p} + \frac{c_p}{c_v}\frac{dv}{v} = 0.$$

Integrating, $$\log p + \frac{c_p}{c_v}\log v = \text{constant}.$$

Calling c_p/c_v, k, we may write $$pv^k = \text{constant}. \tag{141}$$

The general gas law may be applied to the gas before and after the adiabatic compression, giving the three equations $p_1v_1^k = p_2v_2^k$, $p_1v_1 = MrT_1$, and $p_2v_2 = MrT_2$.

PROBLEMS

1. From the three equations above, show that, for an adiabatic change,

$$\left(\frac{v_2}{v_1}\right)^{k-1} = \frac{T_1}{T_2} \text{ and } \left(\frac{p_1}{p_2}\right)^{\frac{k-1}{k}} = \frac{T_1}{T_2}.$$

2. A certain quantity of gas initially at 0° C. is suddenly compressed to half its volume. What temperature will the gas attain? The two specific heats of this gas are .24 and .17 cal. per gram. What would be the final temperature if the gas were originally at 50° C.?

3. Prove that in an adiabatic expansion from volume v_1 to v_2 the work done is

$$\int p\,dv = \frac{p_1v_1 - p_2v_2}{k-1}.$$

Show that $p_1v_1 > p_2v_2$ so that the work done by the gas expanding adiabatically is a positive quantity.

Hint: $$pv^k = p_1v_1^k = p_2v_2^k.$$

4. How much work is obtained in the sudden expansion of 2 grams of air at 0° C. from 20 atmospheres pressure to one atmosphere pressure? The density of the air under standard conditions is 0.00129 gms/cc. The ratio of the specific heat of air at constant pressure to that at constant volume is 1.403.

84. The Isothermal and Adiabatic Elasticities of a Perfect Gas. — In § 21, we obtained an expression for the volume elasticity of any substance. If a change in pressure dp produces a change in volume dv, then Eq. 39 may be written

$$\beta = -v\frac{dp}{dv}. \tag{142}$$

The negative sign is used because an increase in pressure produces a decrease in volume and hence dp/dv is essentially a negative quantity. If the substance being compressed is a perfect gas, we have the further relation, $pv = NRT$. If the changes in volume are produced isothermally, then we obtain from the gas law

$$p\,dv + v\,dp = 0, \quad \text{or} \quad -v\frac{dp}{dv} = p.$$

Comparing this expression with the one above, we see that

$$\text{Isothermal } \beta = p. \tag{143}$$

The isothermal bulk modulus of a perfect gas is thus found to be equal to the existing pressure of the gas.

If the gas is compressed so that no heat is allowed to flow in or out of it, then the relation between p and v is $pv^k = \text{constant}$. Differentiating this expression,

$$kpv^{k-1}\,dv + v^k\,dp = 0, \quad \text{or} \quad -v\frac{dp}{dv} = kp.$$

Hence

$$\text{Adiabatic } \beta = kp. \tag{144}$$

As we have seen in §§ 59 and 60, the adiabatic bulk modulus of a gas comes into play in very rapid compressions such as occur in sound waves.

Thermodynamics

85. The Thermodynamic or Kelvin Work Scale of Temperatures. — It is the purpose of the following discussion to develop a temperature scale, the Thermodynamic or Work Scale, which does not depend upon the behavior of any particular physical property of a substance and which is, in that sense, a truly absolute temperature scale. In scientific work all temperatures must ultimately be expressed on this scale.

86. Cyclic Processes; Reversible and Irreversible Processes. — Any process in which the working substance is carried through various changes and finally brought back to its original condition is a *cyclic process*. *A process is said to be reversible when, if the conditions which caused the process are reversed, every part of the process is repeated in the inverse order.* For example, if we have a gas contained in a perfectly conducting cylinder and subject it to a pressure which is constantly increasing by infinitesimal steps, a compression at constant temperature ensues. If now we reverse the conditions, that is, continually decrease the pressure by

infinitesimal steps, an expansion at constant temperature ensues such that at each pressure the gas occupies the same volume that it did during the compression. This process is indicated by an isothermal curve on the *pv* diagram. It is to be noted that reversible processes can be only approximately attained. The slower the process the more nearly does the reverse path coincide with the direct one. Of course, we cannot change an equilibrium state without disturbing it and thereby departing from the equilibrium curve. For instance, we cannot move on the isothermal from *a* to *c* (Fig. 69). We may approximately keep to the isothermal by taking a succession of infinitesimal steps as from *a* to *b*, waiting each time for the heat generated to flow out and the temperature to drop back to its original value. Only by making the steps infinitesimal can the process be called isothermal. A sudden large increase in pressure will cause a succession of states represented by the path *adc* which differs widely from the isothermal although the gas will gradually cool and settle into the condition given by the point *c*; decreasing the pressure in the same way will bring the gas eventually back to *a*, but by a wide swing such as *cea*, below the isothermal.

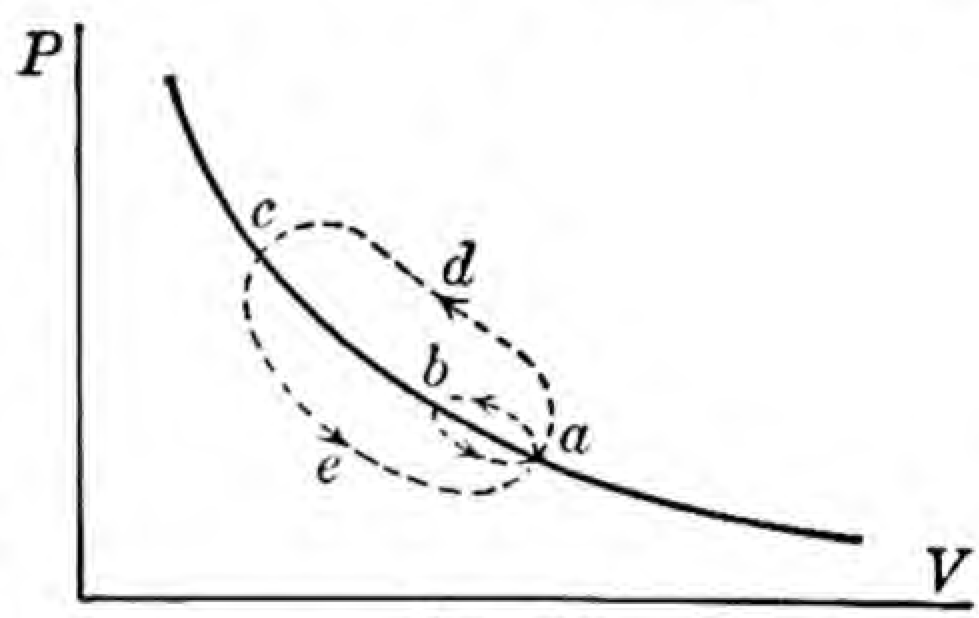


FIG. 69

A reversible process then may be only approximately attained; the more perfectly reversible the process is, the more nearly does it follow along an equilibrium curve.

It should be noted that equilibrium curves are not restricted to isothermals and adiabatics, any curve representing a process carried out at infinitesimal speed would represent a process in which the substance is at all times in equilibrium. For instance, the pressure of a gas might be maintained constant while the temperature is slowly changed causing the gas to expand, or all three variables might be allowed to change. In all cases, if the changes are slow enough, the path may be retraced by merely reversing the conditions.

An outstanding example of an irreversible cycle is any case where friction is present and heat is thereby developed.

87. The Carnot Cycle. — The laws of isothermal and adiabatic changes may be expressed as we have seen in very simple mathe-

matical terms and for this reason a reversible cycle following isothermal and adiabatic paths is from a theoretical standpoint the most convenient and easiest to study. Such a cycle is called a Carnot cycle. A device in which the working substance is carried through such a cycle is called a Carnot engine.

A Carnot engine is only theoretically attainable, but its study has led to many deductions of great practical value to the physicist, engineer, and chemist. Fig. 70 represents a Carnot cycle on a *pv* diagram.

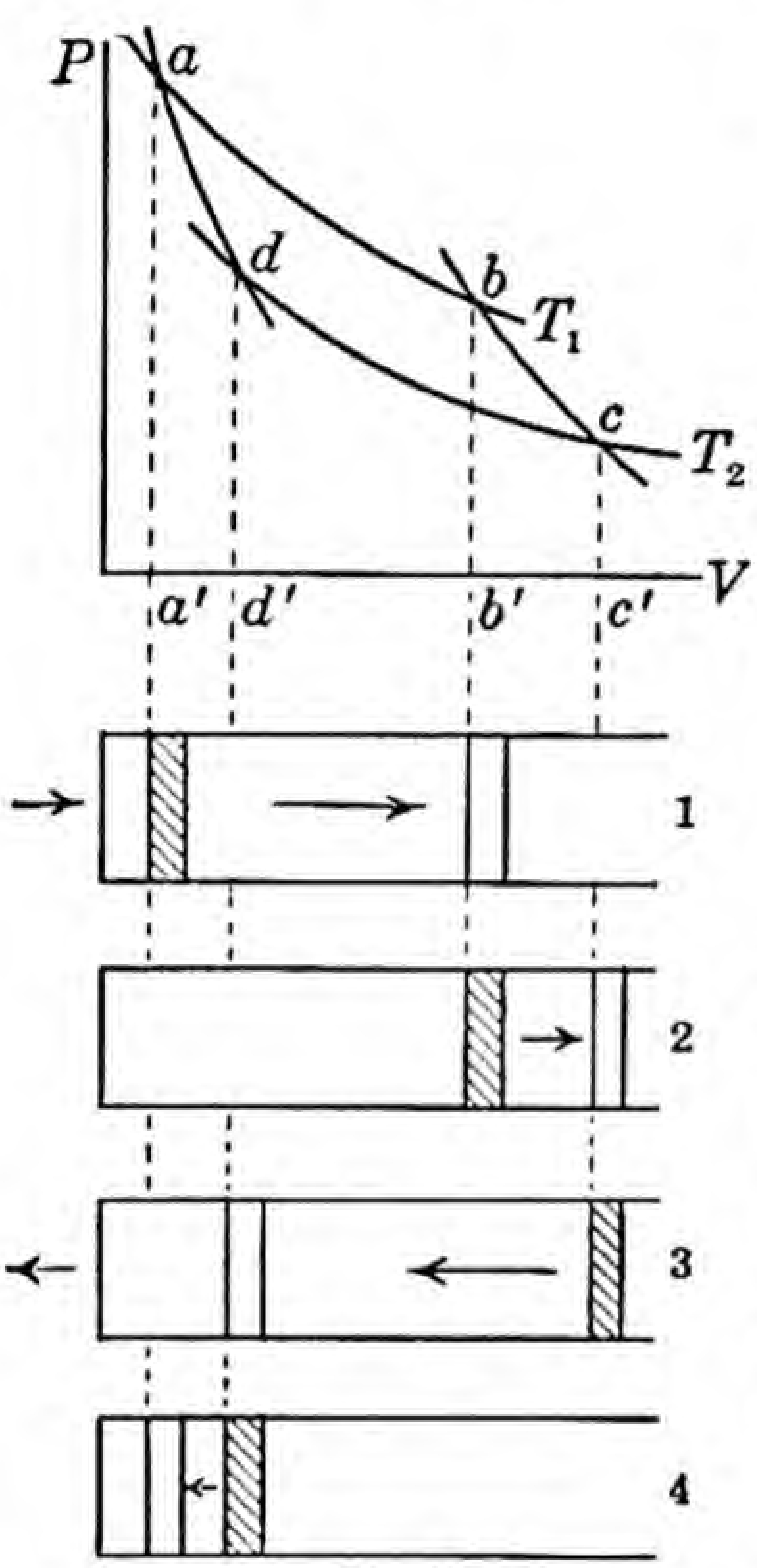


Fig. 70

The gas is taken first at a state represented by point *a*. Consider the gas to be in a cylinder (1) in contact with a source of heat which maintains it at a constant temperature T_1. The gas is allowed to expand slowly to point *b*, heat *flowing in* from the source so that the curve *ab* is an isothermal. The cylinder is then placed so as to be completely heat insulated (2). The gas is allowed to expand adiabatically from *b* to *c*, the temperature dropping to some temperature T_2. The cylinder is then placed on a body that will absorb heat so as to maintain the gas at temperature T_2 (3). The gas is very slowly compressed to some point *d*, *heat flowing out* of the gas so that the curve *cd* is an isothermal. Then the cylinder is again thoroughly insulated (4), and the gas compressed adiabatically to the point *a*, the temperature rising again to T_1. The gas was compressed isothermally in the third step to such a point *d* that the adiabatic curve through *d* passes through point *a*.

88. The Efficiency of a Carnot Engine.—It has been shown that an area under a curve on a *pv* diagram, expressed by $\int pdv$, represents work done during the process indicated by the curve. From *a* to *b* to *c* (Fig. 70) the work done *by the substance* is represented by the area *abcc′a′*. From *c* to *d* to *a* the work done *on the substance* is indicated by the area *cdaa′c′c*. The *net* work obtained during the process is therefore the area *abcd*. Heat Q_1 is added to

the substance during the change ab and heat Q_2 flows out of the substance during the change cd. The net heat disappearing or transformed to work is thus $Q_1 - Q_2$. Thus $Q_1 - Q_2 =$ area $abcd$. The *efficiency* is the ratio of the heat converted into work to the total heat taken in.

$$\text{Efficiency} = \frac{Q_1 - Q_2}{Q_1}. \tag{145}$$

As just seen, the numerator, $Q_1 - Q_2$, is represented by the area $abcd$. The denominator cannot in general be represented by a definite area in the Carnot cycle diagram.* However, in the case of a perfect gas we know that during an isothermal expansion from a to b there can be no change in the internal kinetic energy and under no condition can there be any internal potential energy, so all the heat added, Q_1, must be used to do external work and is represented by the area $abb'a'$. Hence, *for a perfect gas,*

$$\text{Efficiency} = \text{area } abcd \div \text{area } abb'a'.$$

89. The Carnot Cycle Not Limited to Gas Engines. — When any material is carried through a reversible cycle of any sort of isothermal and adiabatic changes, the cycle is a Carnot cycle and the efficiency equation (Eq. 145) holds. The following is an example. Fig. 71 shows the force-elongation relations for a wire at temperature T_1 and T_2. Consider the wire to be in the state represented by the point a. Let us increase the load on the wire very slowly, allowing an inflow of heat to prevent the cooling which the stretching would otherwise entail (isothermal change from a to b). Now let a large load be applied suddenly. The temperature will fall to T_2 (adiabatic change from b to c). Let the surroundings now be at the temperature T_2 and slowly diminish the load, allowing heat to flow out of the wire to prevent a rise in temperature which the shrinking would otherwise entail (isothermal change from c to d). For most substances Young's modulus increases with decreasing temperature as is shown by the difference in slopes of the two

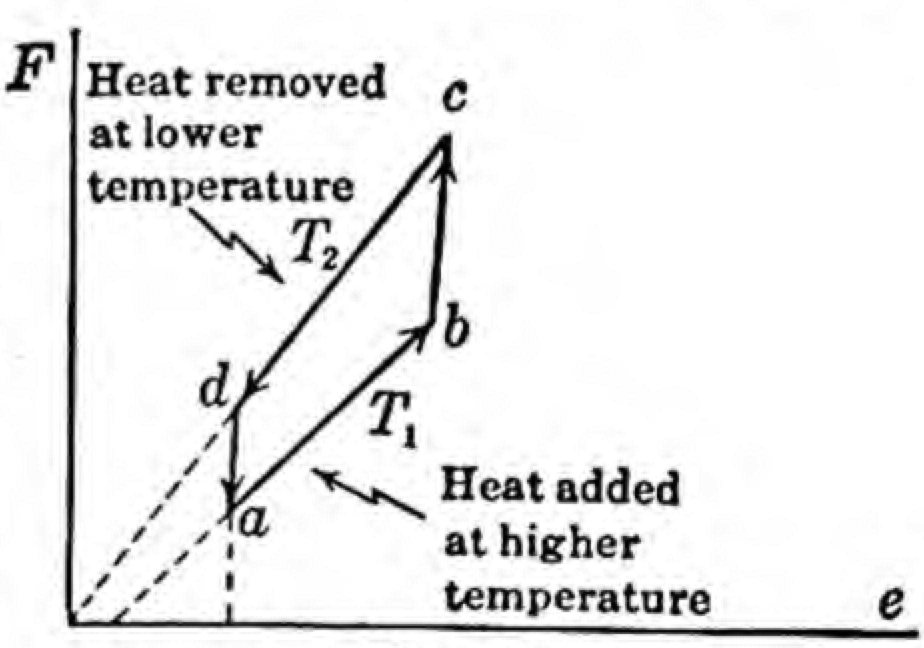


FIG. 71

* See § 101 and Problem 9, § 78, where the work done is a very small part of the heat added.

isothermals. Finally, a large load is removed all at once and the temperature of the wire rises to T_1 (adiabatic change from d to a). Areas under a force-elongation curve represent work. The area under abc gives the amount of work done *on* the system and during this change heat Q_1 is added to the system from the boiler at the higher temperature. The area under cda gives the amount of work done *by* the system and during this change heat Q_2 is removed from the system at the lower temperature. The area $abcd$ represents the net work gained during the cycle, obviously the mechanical equivalent of $(Q_1 - Q_2)$.

90. The Second Law of Thermodynamics. — The First Law of Thermodynamics states "*whenever* heat is changed into work, or work into heat · · · " and does not state anything concerning the conditions under which the changes may be made. The Second Law of Thermodynamics considers the conditions under which heat energy may be converted into work.

We find by experience that some mechanism is always necessary in order that the heat may produce work, that there must be some substance which receives the heat and undergoes some change such as expansion, change of state, etc., which involves the performance of work, and that if the machine is to run continuously of its own accord, it must repeatedly go through some cycle in half of which work is done by the substance and in the other half of which a smaller quantity of work must be done on the substance to get it back into its original condition. As in Carnot's cycle, it is always found that such machines take in, during each cycle, a certain quantity of heat from a body at one temperature and return a smaller quantity of heat to some other body at a lower temperature. The natural tendency is for heat to flow from a hotter to a cooler body and it is this tendency which makes it possible for an engine to operate continuously between two temperatures and do work. On the contrary, whenever heat is taken from a colder body and transferred to a hotter one, as is accomplished in all mechanical refrigerators, work must be continuously done on the machine; in other words, such a refrigerator must be aided by an external agency. Based on the above stated experimental findings, the Second Law may be stated: *It is impossible for a self-acting machine working in a cycle, unaided by an external agency, to convey heat from one body to another body whose temperature is higher; or heat cannot of itself* (*i.e.* without the performance of work by some

external agency) *pass from the colder to a hotter body.* No experimental fact has yet been found that is contrary to this law. So firmly has the law stood every test, that it is now taken as the criterion of future performance. It is firmly believed that no machine can work which violates this law.

91. Carnot's Theorem. — *All reversible engines working between the same temperatures have the same efficiency and no irreversible engine working between the given temperatures can have a greater efficiency than a reversible engine.* If this can be shown, it follows that efficiency depends only upon the temperatures between which the engine works and that the expression for efficiency which we may be able to obtain by means of a Carnot engine will hold for any reversible engine and will express the maximum efficiency attainable between the two temperatures considered. The proposition may be proved in the following way.

Consider two reversible engines E and E' (Fig. 72), each working between the same two temperatures and between the same heat source and the same exhaust tank. Let these engines be mechanically connected, the more efficient one running the less efficient one backward. Necessarily, all the energy output of E is used in running E' backward, so the net work of the two machines is zero. The less efficient engine, E', in order to develop the same horsepower as E, must take more heat from the source and, therefore, pass on to the exhaust tank more heat as waste than the other (indicated by dotted arrows). The less efficient engine is now taking heat Q_2' from the exhaust tank and takes out more heat than the efficient engine puts in. Also the less efficient engine passes into the source the heat Q_1' which is more than the more efficient one takes out. As a net result of this whole process we have heat flowing from a cold to a hot region without any energy being required to transfer it. This is contrary to the experience embodied in the second law and, therefore, we conclude that both such engines must have the same efficiency.

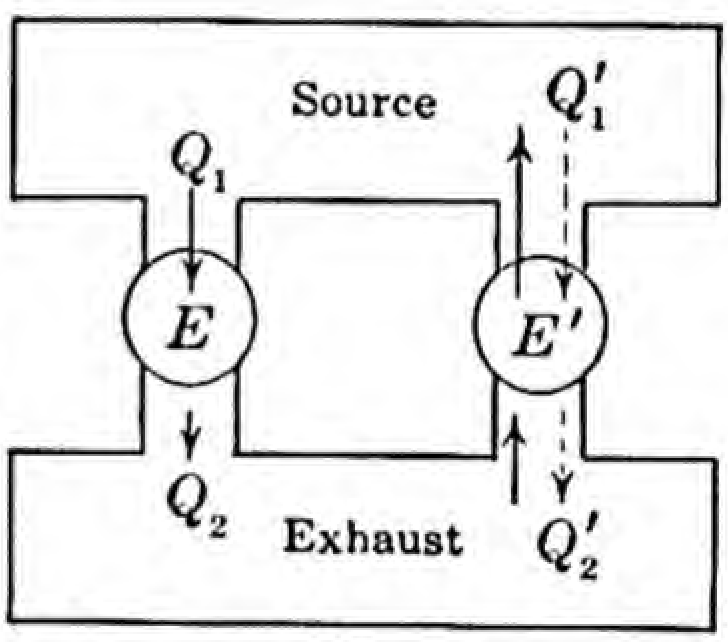


FIG. 72

Suppose an irreversible engine to be more efficient than a reversible engine. Let E be irreversible and E' reversible and let E run the reversible engine E' backward. The result would be identical

to the above case, — we have heat flowing without compensation from a cold to a hot region. Hence no irreversible engine can have a greater efficiency than a reversible engine working between the same temperatures.

Throughout the above proofs it has not been necessary to inquire what type of engine was used or what working substances were used in the engines. It was required only that the engines work in a reversible cycle between two temperatures. Therefore, the efficiency of a reversible engine must depend only upon the temperatures between which the engine operates.

92. The Thermodynamic Temperature Scale. — Since the efficiency of a reversible engine is a function of the temperatures only, we may write

$$\text{Efficiency} = \frac{Q_1 - Q_2}{Q_1} = f(T_1, T_2). \tag{146}$$

Let a Carnot engine operate between a source of fixed temperature T_1 and an exhaust at any other temperature T_2. Let the engine take from the source a fixed amount of heat Q_1. When Q_1 and T_1 are fixed, we see from Eq. 146 that the work $Q_1 - Q_2$ obtained from the engine is a function of the temperature T_2 only. Therefore the output of a Carnot engine is certainly a thermometric property of the engine and we may establish a temperature scale based upon it. Temperatures on this new scale will be indicated by the letter θ.

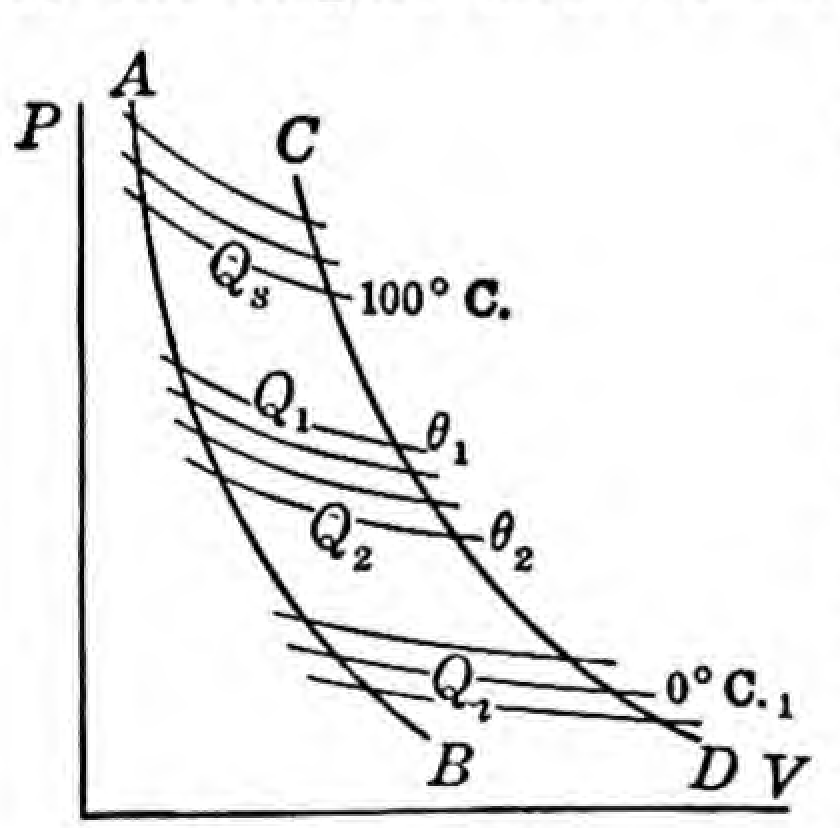


Fig. 73

Consider an engine operating between 100° and 0° C. and along the two adiabatic curves AB and CD (Fig. 73). Let Q_s be the quantity of heat taken in from the source at 100° C. and Q_i be that delivered to the exhaust at 0° C. Consider the area between the 100° and 0° C. isothermal lines to be divided into 100 equal areas by 99 intermediate isothermal lines. The temperature difference indicated by any adjacent two of these isothermals is defined to be a one-degree interval on the thermodynamic scale. An engine operating between a source and exhaust whose temperature difference is one such degree will have an output of $(Q_s - Q_i)/100$.

Now let an engine operate from a source at a temperature of θ_1, taking in heat Q_1, and giving out heat Q_2 to the exhaust at a tem-

perature of θ_2. By our above definition we have for the temperature interval on the thermodynamic scale, the expression

$$\theta_1 - \theta_2 = \frac{Q_1 - Q_2}{Q_s - Q_i} 100. \tag{147}$$

On this new temperature scale $\theta_1 - \theta_2$ is always directly proportional to the output $Q_1 - Q_2$. We may now rewrite Eq. 146, using the above ideas.

$$\text{Efficiency} = \frac{Q_1 - Q_2}{Q_1} = \frac{k(\theta_1 - \theta_2)}{Q_1}, \tag{148}$$

where k is a proportionality constant.

We must now find the value of the factor k/Q_1. Imagine that the temperature of the engine exhaust is lowered degree by degree. At each step a further portion of the energy Q_1 is transferred into work,—an amount equal to $(Q_s - Q_{i1})/100$ for every degree drop. Now, obviously, this procedure will reach a limit when a temperature has been reached at which all the heat has been transferred into work. No lower temperature than this is conceivable because, according to experiment, out of a given quantity of heat only a certain amount of work may be obtained (the First Law of Thermodynamics). This point on the thermodynamic scale, where all the heat is convertible into work, must be the absolute zero. With an engine operating between any temperature θ_1 and the absolute zero the efficiency must be unity. Hence Eq. 148 gives

$$\frac{k}{Q_1} = \frac{1}{\theta_1}.$$

Therefore for any engine operating between the temperatures θ_1 and θ_2,

$$\text{Efficiency} = \frac{\theta_1 - \theta_2}{\theta_1}. \tag{149}$$

This temperature scale is not peculiar to the particular engine chosen. Any other reversible engine would yield the same scale. For let two reversible engines operate between a given source and exhaust. Let one engine indicate the temperatures to be θ_1 and θ_2 while the other engine indicates them to be θ_1' and θ_2'. By Carnot's theorem these engines must have the same efficiency, so

$$\frac{\theta_1 - \theta_2}{\theta_1} = \frac{\theta_1' - \theta_2'}{\theta_1'}. \tag{150}$$

Let the two engines operate between the steam and ice points. Then $\theta_1 - \theta_2 = \theta_1' - \theta_2' = 100°$ (or $180°$ for a Fahrenheit scale)

and by Eq. 150, $\theta_1 = \theta_1'$. This means that if we choose the same temperature scale (either Fahrenheit or Centigrade) for both engines, then both engines must give the same reading for the steam point.

Now operate the two engines between the steam point and any arbitrarily chosen temperature. Since $\theta_1 = \theta_1'$, then

$$\frac{\theta_1 - \theta_2}{\theta_1} = \frac{\theta_1 - \theta_2'}{\theta_1}.$$

Hence $$\theta_2 = \theta_2'.$$

Since θ_2 is any arbitrary temperature, the two Carnot engines must agree at all temperatures. Therefore we have proved that if the same scale is selected for any two Carnot engines whatsoever, the temperatures indicated by those engines will always be the same.

The value of the thermodynamic temperature scale lies in the fact that it is a truly absolute scale in the sense that it does not depend upon the behavior of any physical property of a substance. It is, therefore, chosen as the standard scale, all others being referred ultimately to it (§ 98).

93. The Maximum Efficiency of a Heat Engine. — It has been shown that no engine working between two given temperatures can have a greater thermal efficiency than a Carnot engine and that this maximum efficiency is given by the expression $(\theta_1 - \theta_2)/\theta_1$. The upper limit of possible efficiency is thus fixed for any heat engine. It should be noted that in this expression for efficiency temperatures must be expressed on the thermodynamic scale.

94. Entropy. — If a substance is carried around a Carnot cycle we may write the efficiency equation as

$$\frac{Q_1 - Q_2}{Q_1} = \frac{\theta_1 - \theta_2}{\theta_1}.$$

This may be expressed in the form

$$\frac{Q_1}{\theta_1} = \frac{Q_2}{\theta_2}. \tag{151}$$

The quantity $$Q_2 = \frac{Q_1}{\theta_1}\theta_2$$

represents that amount of heat which is given to the exhaust chamber and cannot be used by the engine. It is completely unavailable as a source of work. If the exhaust temperature θ_2 is held constant while θ_1 is varied, it is seen from the above

expression that Q_1/θ_1 is the factor which determines the variation in the amount of unavailable energy, Q_2. No matter what quantity of heat Q_1 enters the engine from the source, the fractional part $1/\theta_1$ of it will be unavailable for each degree Absolute the exhaust temperature is above the absolute zero. This quantity Q_1/θ_1 is called the *change in entropy* of the body which the heat leaves or enters. When a quantity of heat Q is added to a body at temperature θ, Q/θ is the increase in entropy of that body. When a quantity of heat Q passes out of a body whose temperature is θ, then the entropy of that body is decreased by Q/θ. So we see from Eq. 151 that for a Carnot cycle the change in entropy of the system (boiler and condenser) is zero since the entropy of the exhaust is increased by the same amount as the entropy of the boiler is decreased.

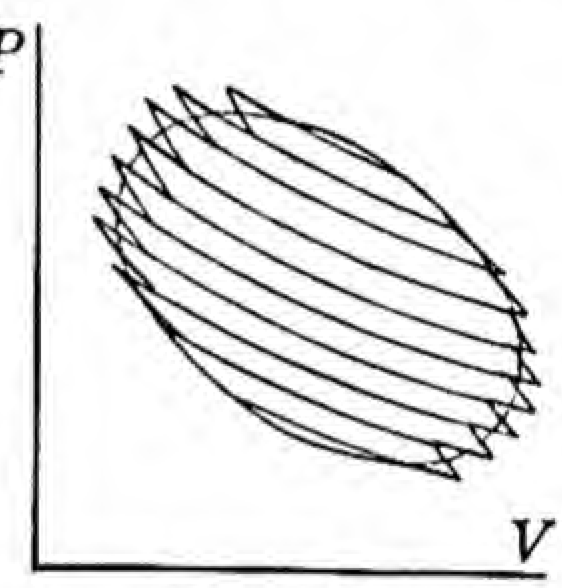


FIG. 74

It then follows that the entropy change for *any* reversible cycle is zero, because any reversible process may be broken up into small isothermal and adiabatic steps and any reversible cycle considered as the sum of a large number of infinitesimal Carnot cycles (Fig. 74). For each small cycle the entropy change is zero, and for the complete cycle considered as made up of infinitesimal Carnot cycles the total change is zero or

$$\sum \frac{dQ}{\theta} = 0 \tag{152}$$

for any reversible cycle.

For an irreversible cycle

$$\frac{Q_1 - Q_2}{Q_1} < \frac{\theta_1 - \theta_2}{\theta_1},$$

that is, the efficiency is less than that for a reversible cycle between the same temperatures. From this it follows that $Q_2/\theta_2 > Q_1/\theta_1$. So in any irreversible cycle there is an increase in the entropy of the system.

If the temperature of the system changes while the heat is being added, then the change in entropy is

$$\int_{\theta_1}^{\theta_2} \frac{dQ}{\theta}.$$

In almost all practical machinery, 0° C. is the lowest exhaust temperature and so all engineering tables give the entropy of water or

steam with reference to water at 0° C. as being at zero of entropy, just as we measure heights above sea level, taking sea level as a reference zero.

Eq. 152 is sometimes taken as a mathematical statement of the Second Law of Thermodynamics. It says that the entropy change for a reversible cycle is zero. Since no process can be actually reversible and since friction and dissipation of heat occur with every process known, it follows that *every actual process in nature is accompanied by an increase in entropy.*

Every physical process involves an increase in entropy either through heat conduction or the conversion of potential or kinetic energy or work to heat. In the conduction process the heat dQ still exists as heat but at a lower temperature. There has been an increase in entropy. In the conversion of work to heat there has been brought into existence an amount of entropy, dQ/θ, which did not exist before. No process in nature can result in a decrease in entropy, the best that can be done is to keep the increase as small as possible by making our processes as nearly reversible as possible.

These continual increases in entropy mean continual decreases in the supply of energy available for work and continual increases in the energy no longer useful or available for work. It means a degradation toward some common temperature level for all the energy of the universe. When this common level is reached, no further change or physical process will be possible.

95. The Temperature-Entropy Diagram. — If entropy, S, is plotted as abscissa against temperature as ordinates, relationships are much more evident than on the corresponding p–v diagram.

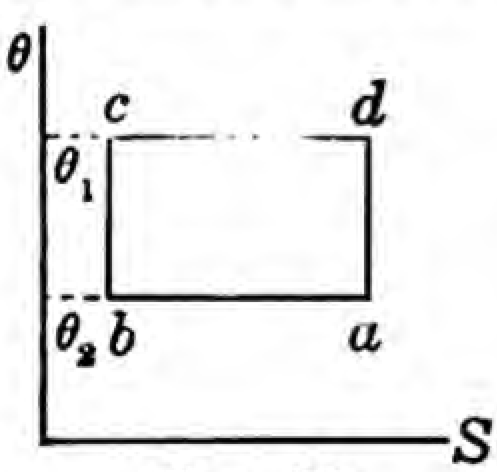


FIG. 75

Consider the cycle as in Fig. 70 or 76. Along the path cd, heat is added to the system and the entropy increases. This is represented in the θ–S diagram as cd (Fig. 75), a straight line parallel to the S axis. Along the adiabatic da, no heat is added, so dQ and dS are zero although θ changes from θ_1 to θ_2. Adiabatic curves on the θ–S diagram are thus lines parallel to the θ axis. From a to b the entropy of the system decreases and from b to c the entropy remains unchanged although the temperature increases to θ_1 again.

$$dQ = \theta\, dS \quad \text{or} \quad Q = \int \theta\, dS.$$

Therefore, since Q has the dimensions of energy (or work) we see that areas on the θ–S diagram represent work just as those on the p–v diagram (or calories, if S is in calories per degree).

96. Comparison of the Perfect Gas Scale with the Thermodynamic Scale. — It will now be shown that the efficiency of a Carnot engine is also given by the equation

$$\text{Efficiency} = \frac{T_1 - T_2}{T_1},$$

where T_1 and T_2 are temperatures indicated on the perfect gas scale. This having been shown, it follows that $\theta_1/\theta_2 = T_1/T_2$. This equality states that the temperatures on the two scales are proportional to each other. It is left as an exercise for the student to prove that $\theta_1 = T_1$ and $\theta_2 = T_2$, or that the perfect gas scale is identical with the thermodynamic scale.

The efficiency of a Carnot engine working on the cycle a, b, c, d (Fig. 76) between temperatures (indicated on the perfect gas scale) T_1 and T_2 is

$$\text{Efficiency} = \frac{Q_1 - Q_2}{Q_1}.$$

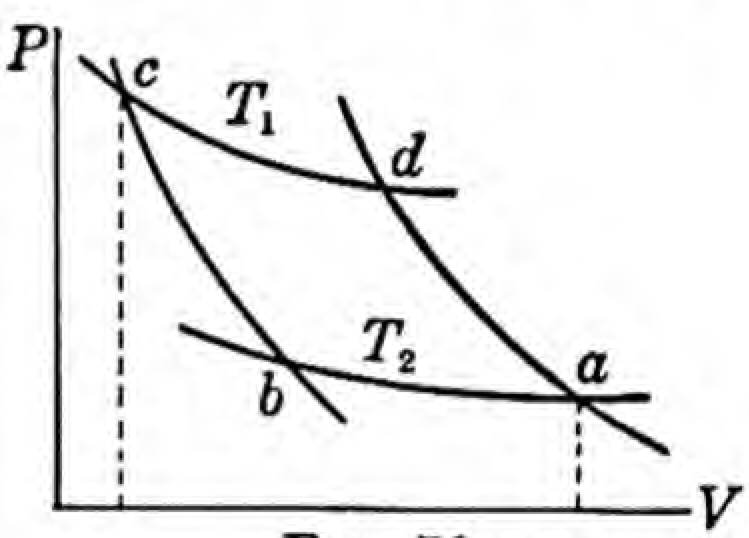


FIG. 76

The heat quantities Q_1 and Q_2 are transferred while the gas is not changing temperature. So in Eq. 138, $dT = 0$ and for a perfect gas we also have $U_p = 0$. We then have

$$Q_1 = \int_c^d p\,dv = NRT_1\int_c^d \frac{dv}{v} = NRT_1 \log_e \frac{v_d}{v_c},$$

$$Q_2 = \int_b^a p\,dv = NRT_2\int_b^a \frac{dv}{v} = NRT_2 \log_e \frac{v_a}{v_b}.$$

Hence
$$\text{Efficiency} = \frac{T_1 \log_e \frac{v_d}{v_c} - T_2 \log_e \frac{v_a}{v_b}}{T_1 \log_e \frac{v_d}{v_c}}.$$

Since the cycle is made up of two isothermals and two adiabatics we have the equations,

$$p_a v_a = p_b v_b, \quad p_b v_b{}^k = p_c v_c{}^k,$$
$$p_c v_c = p_d v_d, \quad p_d v_d{}^k = p_a v_a{}^k.$$

Multiplying these four equations together and cancelling, we find that

$$\frac{v_d}{v_c} = \frac{v_a}{v_b}.$$

Hence
$$\text{Efficiency} = \frac{T_1 - T_2}{T_1}. \tag{153}$$

By an argument similar to that in the latter part of § 92 we may prove that $\theta_1 = T_1$, and $\theta_2 = T_2$.

PROBLEMS

1. What is the change in entropy of 5 gms. of ice when sufficient heat is added to melt it?

2. An engine works between temperatures of 110° C. and 20° C. and its efficiency is 20 per cent. What is the change in entropy of the system for each 100 calories supplied at the boiler?

3. Show that if the specific heat of a solid body remains constant, the change in entropy in heating M grams of it from T_0 to T is

$$Mc \log_e \frac{T}{T_0}.$$

4. In the case of water, where the specific heat differs from unity, we may write $c = 1 + k$, where k is a small quantity having positive values from 0° C. to 15° C. and from 75° C. to 100° C. and negative values from 15° C. to 75° C. Show that the increase in heat content of water from 0° C. is given by

$$\frac{Q}{M} = t + \int_0^t k\, dt.$$

In what units will Q be given? Show how this may be evaluated graphically, given a curve showing the variation of the specific heat with temperature.

5. Show that the entropy of a unit mass of water at $T°$ (measured from 0° C.) is

$$S = \log_e \frac{T}{273} + \int_0^t k\,\frac{dt}{T}.$$

and show how it may be evaluated graphically.

6. Show that if the specific heats of a gas remain constant and if as heat is added to a gas it is allowed to expand, then the increase in entropy is

$$\frac{S}{M} = c_v \log_e \frac{T_2}{T_1} + \frac{N}{M} R \log_e \frac{v_2}{v_1}.$$

7. What is the greatest possible efficiency of an engine working between the temperatures 140° C. and 20° C.?

8. What is the advantage of superheating steam in the case of a steam engine? of using a condenser? What possible advantages might accrue from using mercury as a working substance in place of water? Such mercury vapor engines are actually in operation.

9. An ideal engine whose condenser temperature is 30° C. has an efficiency of 30 per cent. It is desired to raise the efficiency to 40 per cent. What increase in the temperature of the boiler would produce the change? What change in the temperature of the condenser would produce the change?

10. Find the reading of absolute zero on the Fahrenheit scale.

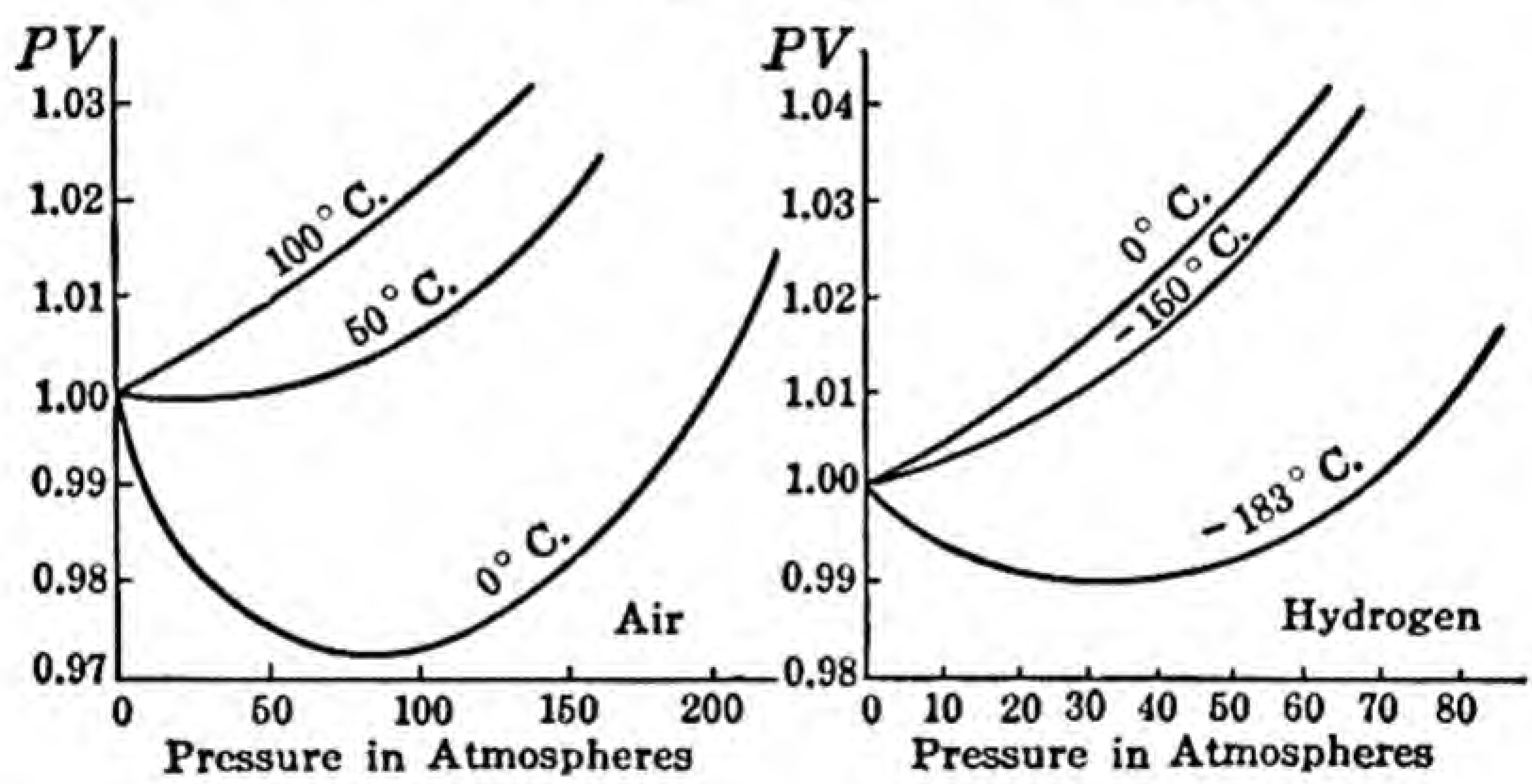
11. An ideal engine receives 100 B.T.U.'s at 350° F. and rejects 70 B.T.U.'s to the exhaust. What is the temperature of the exhaust?

12. An ideal engine operates with its condenser at 40° F. It does 300 ft. lbs. of work for every 3 B.T.U.'s delivered from the boiler. What is the temperature of the boiler?

13. Give the algebraic proof of the first half of § 91, starting with

$$\frac{Q_1' - Q_2'}{Q_1'} < \frac{Q_1 - Q_2}{Q_1}.$$

97. Deviations from Boyle's Law. — (1) *Amagat's Curves.* — According to Boyle's law the product of the pressure and the volume of a gas should be constant if the temperature of the gas is kept constant. In other words, the product pv should not change as p is changed provided T is not changed. The first accurate measurements of the value of pv over a wide range of pressure were



(a) (b)

FIG. 77

made by Amagat in 1893. As shown in Fig. 77, deviations from Boyle's law at high pressures are very marked. The value of pv in the case of hydrogen at 0° C. continually increases while for air at 0° C. the value of pv at first decreases but at higher pressures increases. Experiments of more recent years show that when hydrogen is cooled to − 183° C. a curve (see Fig. 77*b*) is obtained very similar to that for air at 0° C. On the other hand, if air is heated to 100° C. its curve becomes similar to that of hydrogen at 0° C. It is now known that for every gas there is a certain temperature above which the values of pv increase with increase in

p and below which the value of pv at first decreases and then increases.

(2) *The Joule-Thomson Experiment.* — Consider a cylinder (Fig. 78) into one end of which a gas is being pumped so that a constant pressure p_1 results. Let this gas pass through a plug P of porous material so that the gas expands and emerges at a constant pressure p_2. The purpose of the porous plug is to let the gas expand and yet not let the molecules of the emerging gas acquire any considerable forward velocity, which they would acquire if the gas expanded through a single nozzle. If the expansion is carried out so that $p_1v_1 = p_2v_2$, no net external work is done by the gas. Such a process is called *free expansion.* The free expansion of a perfect gas should produce no change in the temperature. However, for all real gases there is a temperature change. At room temperature most gases cool, but hydrogen and helium (and doubtless neon) are warmed slightly upon undergoing free expansion. It is now known that for expansion from a given pressure down to atmospheric pressure there are for certain gases (and probably for all) certain temperatures above which free expansion will cause heating and below which cooling will result. These inversion temperatures, although intimately connected with, are not to be confused with those at which the changes occur in the Amagat curves. For example, below $-80°$ C. free expansion of hydrogen from 113 atmospheres produces cooling while the change in the Amagat curves does not occur until about $-180°$ C.

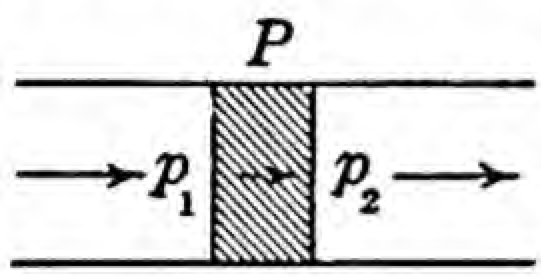


FIG. 78

The original experiments of Joule were carried out to test for the existence of internal attractive forces between the molecules of a gas. If intermolecular forces exist, the gas undergoing the free expansion must supply of itself the increase in the potential energy of the separated molecules. Thus the cooling would indicate work against molecular attraction. However, the phenomena which show up in the porous plug experiment are due to both molecular attraction and molecular size. Van der Waals' equation considers both of these facts and a study of the equation shows that if a gas followed that law, there should be such an inversion temperature as described above. In actual performance of the porous plug experiment, it is impossible to keep $p_1v_1 = p_2v_2$ and therefore there results a further change in temperature caused

by the external work of expansion, and corrections have to be made for that effect.

98. Deviations of Gas Thermometers from the Thermodynamic Scale. — We cannot build a Carnot engine and we have no perfect gas; hence, we have no direct method of determining a true thermodynamic temperature. We can, however, determine to what extent any given gas differs from the perfect gas law, $pv = RT$. A perfect gas exhibits no molecular attractive forces, that is, the internal work of free expansion is zero, and its specific heat at constant pressure is independent of the temperature. Hence, by studying pressure and volume relations at various temperatures, by performing the "porous plug" experiment, and by measuring the specific heat at constant pressure for various temperatures, we may determine to what extent a gas differs from a perfect gas. (For further particulars one should consult Poynting and Thomson's Heat and other special treatises.)

The actual corrections for the hydrogen and nitrogen constant volume gas thermometers (in instruments in which $P_0 = 100$ mm. Hg.) required for reduction to the thermodynamic scale are given below. The deviations of the hydrogen scale for practically all purposes are negligible. At high temperatures hydrogen escapes through the walls of most containers, so other gases must be used.

Temperature Centigrade	Helium	Hydrogen	Nitrogen
− 250	+ 0.04	+ 0.12	
− 150	+ 0.01	+ 0.03	+ 0.2
− 50	+ 0.002	+ 0.005	+ 0.03
0	0.000	0.000	0.000
+ 25	− 0.001	− 0.001	− 0.008
+ 50	− 0.001	− 0.002	− 0.010
+ 75	− 0.001	− 0.001	− 0.005
+ 100	0.000	0.000	0.000
+ 200	+ 0.006	+ 0.02	+ 0.02
+ 400	+ 0.04		+ 0.14
+ 800			+ 0.5
+ 1200			+ 1.0

99. Liquefaction of Gases. — Gases whose critical temperatures are above room temperature may be liquefied at room temperature

by compression alone. The critical temperatures of oxygen, hydrogen, nitrogen, and most gases are so low that special means have to be used to reduce their temperatures below that point. This is brought about by making use of the cooling of the gas by expansion. This cooling, as explained before, will be due to the Joule-Thomson effect and to external work done during expansion.

In the Linde or Hampson process the gas, thoroughly freed from water vapor, is compressed in C (Fig. 79) and then cooled in an ice-water bath W. The cool air then passes into an expansion chamber E, escaping from a jet regulated by R. Cooling results from the Joule-Thomson effect. The air thus cooled is directed back through the expansion chamber, cooling the incoming gas still more. As this process continues, the gas entering the expansion jet becomes colder and colder until it gets sufficiently below its critical temperature to liquefy under the pressure maintained by the compressor. From that point on, part of the expanding air is liquefied in the jet and drops through an opening into a Dewar vacuum flask.

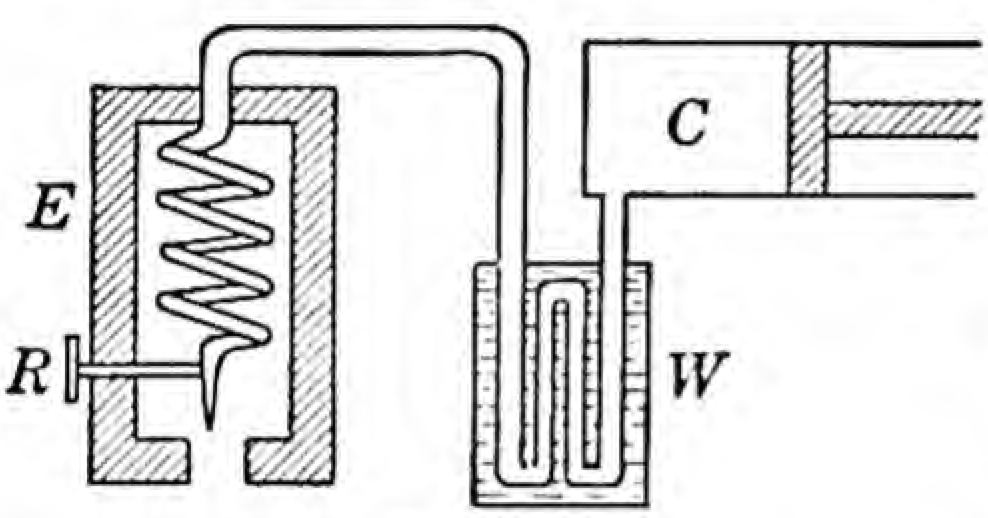


FIG. 79

In the Claude method higher efficiency of production is attained by letting the gas, as it expands from the jet, do work on another piston. The energy for that work must come from the expanding gas so that the gas is cooled more than it would be otherwise. The amount of work gotten out of the second piston may be used to help compress the incoming gas, but since that is only a few per cent of the total energy required, the saving is not large. However, the process of liquefaction of the gas is so aided by the extra cooling thus obtained that the quantity of liquid gas produced, for a given amount of energy delivered to the compressor, is greatly increased.

As mentioned before, hydrogen at ordinary temperatures heats upon free expansion. However, when the compressed hydrogen is cooled to liquid air temperature, it then cools upon expanding and thus may be liquefied.

100. Mechanical Refrigeration. — As mentioned in § 90, it is possible to transfer heat from a cold body to a hotter body by an expenditure of a certain amount of work. In all mechanical re-

frigerators some certain gas is used which will remain in liquid form at room temperature under a moderate pressure, *i.e.* a gas whose critical temperature is not far above room temperature. The most common gases used are ammonia and sulfur dioxide although many others may be used. The gas is compressed in A (Fig. 80) to a high pressure such that by passing it through a radiator R and cooling it either by a current of air or water, the gas liquefies. This step is one in which heat is given out to the hotter body. The liquid gas is piped to a jet from which it escapes into a chamber C which is kept at a low pressure by the suction stroke of the compressor. The liquid upon entering C first evaporates and then expands greatly. The energy necessary to separate the molecules of the liquid against the strong molecular attraction and to do the work of expansion against the vapor pressure in C must be supplied from the kinetic energy of vibration of the molecules and so a great reduction of temperature results. In small refrigerators C is a long coil of copper tubing in the top of the cooling chamber. Heat flows from the chamber into this cold coil. This expanded gas, therefore, becomes warmer as it reaches A, where it is compressed again and the cycle repeated. Heat is continuously drawn from the refrigerator and given out at the radiator at the expense of mechanical work done on the working substance by the compressor. In larger refrigerating systems, C is immersed in brine which is in continuous circulation through insulated pipes to distant refrigerator rooms.

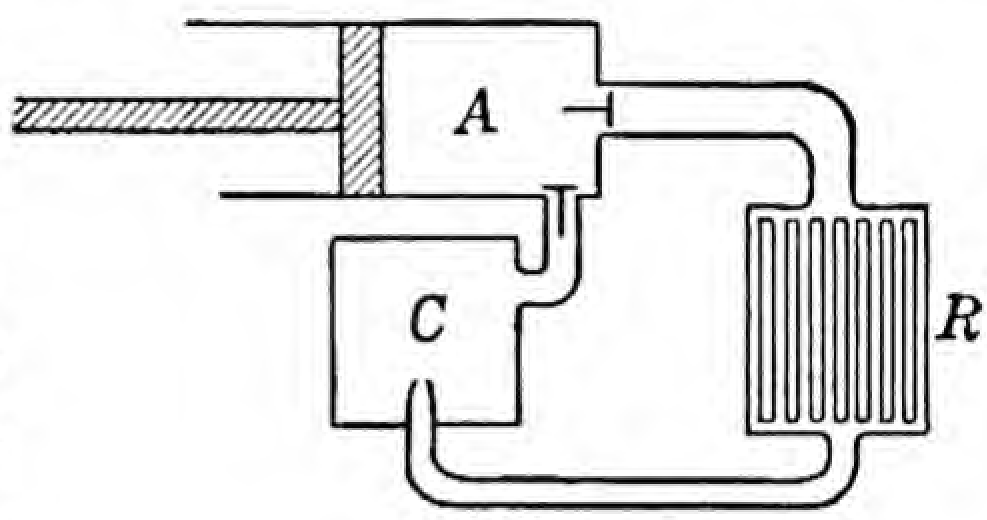


FIG. 80

It is to be noted that very little cooling results from the work of expansion against the attractive forces between the molecules of the vapor in expanding, most of it coming from the change in state and expansion against external pressure.

101. The Use of the Carnot Cycle in Deriving Thermodynamic Equations. — The method of reasoning in which a working substance is carried through a Carnot cycle and in which the relations established for such a cycle are made use of, that is, thermodynamic reasoning, has led to the establishment of many relations of theoretical and practical importance. As an example of this method, suppose it is desired to find the rate of change with temperature of

the pressure of a saturated vapor. Take two isothermal curves, $ABCD$ and $A'B'C'D'$, such as described in § 87, for two very slightly different temperatures (Fig. 81). Let the curves be drawn for one gram of the substance, so that v_2 is the volume of one gram of the saturated vapor at temperature θ_2 and v_1 the volume of one gram of the liquid. At C' the vapor is saturated at the temperature θ_2. At the point B the gram of matter is all liquid. As the

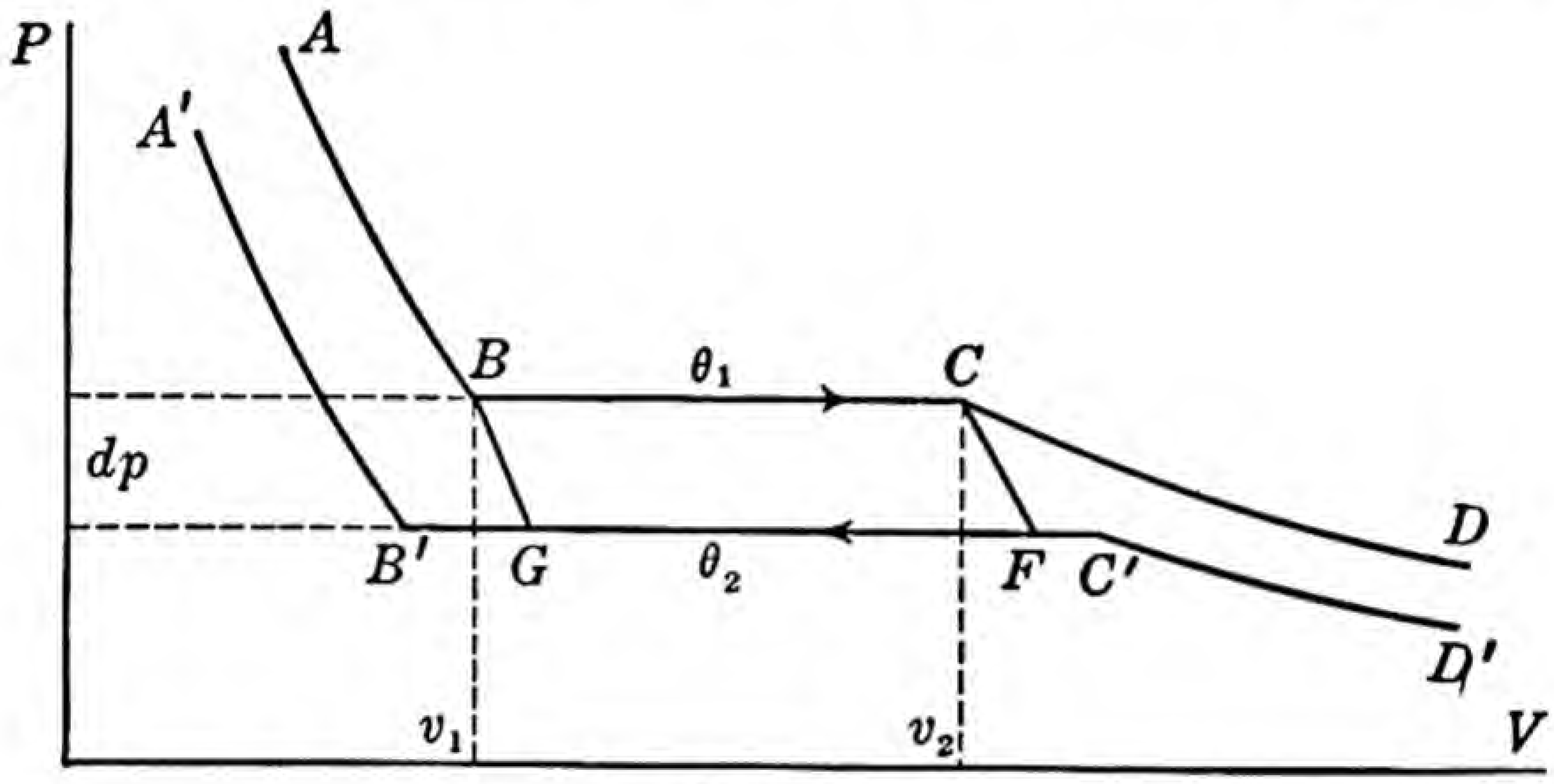


FIG. 81

volume is increased, some of the liquid vaporizes and heat Q_1 has to be added to maintain the temperature θ_1 constant. The pressure remains constant since the pressure of a saturated vapor cannot change unless the temperature changes. When all the liquid has evaporated, the vapor occupies the volume v_2. The condition of the substance is represented by the point C. Let the vapor expand adiabatically to point F until the temperature drops an amount $d\theta$ to θ_2. For steam, such an expansion causes such a drop in temperature that some of the vapor condenses. Therefore, the point on the line GC' which is reached is a point F to the left of C'. Then the vapor is compressed. As was stated in § 77, the pressure of a saturated vapor must remain constant as long as the temperature remains constant. Hence the compression results in condensation of some of the vapor and the heat of condensation must be removed in order to keep the temperature from rising. Let Q_2 be the heat removed in order to maintain the temperature constant at θ_2. The compression is continued to a certain point G where an adiabatic compression restores the original conditions.

The distance between the lines BC and GF is very small compared with their lengths so that, although the two adiabatic curves

GF and BG are not quite parallel, we may say that the area of $BCFG$ is $(v_2 - v_1)\,dp$.

$$\text{Efficiency} = \frac{Q_1 - Q_2}{Q_1}$$
$$= \frac{\text{Area } BCFG}{Q_1} = \frac{\theta_1 - \theta_2}{\theta_1} = \frac{d\theta}{\theta_1}.$$

Since $Q_1 = L_v$ the latent heat of vaporization,

$$\frac{(v_2 - v_1)\,dp}{L_v} = \frac{d\theta}{\theta_1}, \tag{154}$$

or

$$\frac{dp}{d\theta} = \frac{L_v}{(v_2 - v_1)\theta_1}. \tag{155}$$

In Eq. 154 it is seen that in order to keep the units correct L_v must be expressed in work units. $v_2 > v_1$ and, therefore, $dp/d\theta$ is positive; *i.e.* $d\theta$ is positive when dp is positive or the boiling point of water is raised when the pressure is raised.

Eq. 155 enables us to compute the value of latent heat of vaporization at any temperature θ provided the change in volume at that temperature and the slope of the vapor pressure-temperature curve is known. At 100° C. one gram of steam occupies 1671 cm³.

$p_{100} = 76.000$ cm. of mercury. $\qquad dp = 0.272$ cm. of mercury.
$p_{100.1} = 76.272$ cm. of mercury. $\qquad d\theta = 0.1°$.
$\theta = 373°$.

$$L_v = \frac{0.272 \times 13.6 \times 981 \times 1670 \times 373}{0.1} \text{ ergs} = 540.4 \text{ calories.}$$

The measured value of L_v is 539.5 calories. The per cent of difference between these two is no larger than the per cent error in the experimental value of dp.

In a like manner we might draw the isothermal curves for ice and water. The slopes of these curves would be different from those for water and steam but a similar diagram would be obtained. Under the conditions represented by the point B (Fig. 82), there exists one gram of water. Expand the water isothermally. To do that, heat is taken out of the water until at C all of it is ice.

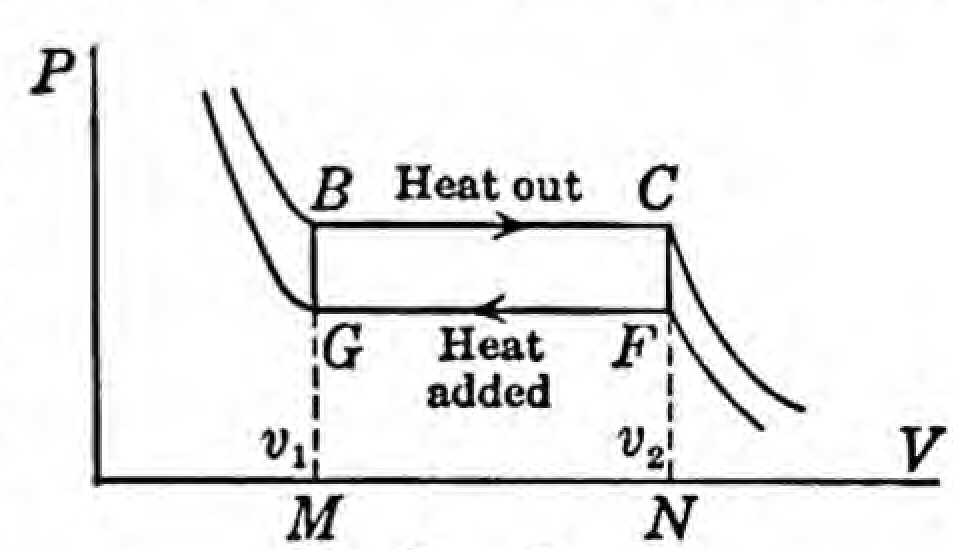


FIG. 82

Then the pressure is reduced by an amount dp so that the ice expands adiabatically to point F. Since very little expansion

occurs, the curve is almost vertical. The temperature change $d\theta$ which occurs will be examined later. Then the ice is compressed isothermally and heat is added while the ice melts until a point G is reached such that an adiabatic compression restores the original conditions. Now suppose that the change F to G occurred at a temperature lower than the original. It is seen that during the expansion an amount of work was done by the gas indicated by the area $BCNM$ while on compression only the work $GFNM$ was required. Therefore, an amount of work $BCFG$ was done by the ice engine during the cycle. But now if FG is at a temperature lower than that for BC, we have an engine operating in a cycle, taking a certain quantity of heat from a colder body, delivering a larger quantity of heat to a hot body, and doing work besides, which is a flat contradiction of the Second Law of Thermodynamics. Therefore, the ice must have heated during the adiabatic expansion. So as the pressure decreases, the temperature increases and $dp/d\theta$ is negative. Now Eq. 155 holds for this case also and so we see that if $L_v/(v_2 - v_1)\theta_1$ is negative, the interpretation is that the latent heat of fusion L_f is taken out of the water while the volume increases from v_1 to v_2, while in the case first discussed L_v was added during the expansion.

In the same fashion as before, we can obtain values of L_f if $d\theta/dp$ is known. However, the data for the change of the melting point with pressure are not easily obtained since the effect is small even under very great pressures. So we had better compute $d\theta/dp$ from the other measured quantities. For this case:

$$\theta = 273° \text{ Ab.}; \; L_f = 334.8 \text{ joules}; \; v_2 - v_1 = 0.0907 \frac{\text{cm}^3}{\text{gm.}}.$$

So

$$\frac{d\theta}{dp} = \frac{0.0907 \times 273}{334.8 \times 10^{-6} \times 1{,}013{,}000} = 0.0073 \text{ degrees per atmosphere.}$$

The best data obtained for this variation place the decrease in the melting point at about 0.0075 degrees per atmosphere.

It is interesting to note that James Thomson, by reasoning similar to that just given, deduced the fact that the melting point of ice should be lowered with increased pressure before the fact was known. Later experiments proved the reasoning to be correct.

PROBLEM

Using the data below, compute the value of L_v from Eq. 155 and compare with the value given. In each case could the difference in the values of L_v be due to an error of 1 or 2 units in the last digit of the vapor pressure? Explain from a molecular viewpoint why the latent heat of vaporization should increase as the temperature decreases. Why should v_2 be so large at 30° C. and so small at 150° C.?

t° C.	p in mm. of Hg.	v_1 in cm³.	v_2 in cm³.	L_v in Joules
30.0	31.824	0.996	32,950	2425
30.1	32.007			
50.0	92.51	0.988	1,202	2379
50.1	92.97			
150.0	3568.7	1.088	392	2107
150.1	3578.2			

The Transfer of Heat

102. Methods of Transfer. — Heat energy may be transferred from one body to another by *conduction* or *radiation*. In conduction the kinetic energy of the more rapidly moving molecules in one region of a body is passed on to the slower moving neighboring molecules by the process of molecular impacts. In radiation the energy is transferred by a wave motion of the same nature as light. These electromagnetic waves are produced partly by the vibration of electrons about the atoms of the body and partly by vibrations of atoms in their molecular groups. Conduction can, therefore, take place only through matter (liquid, solid, or gaseous). Radiation may take place through a vacuum or through transparent or translucent bodies. In passing through any matter, however, some part of the energy of the wave motion is absorbed, *i.e.* transferred to energy of vibration of the molecules.

Molecular kinetic energy may also be transferred from one position to another by transferring the matter possessing the energy. This process is called *convection* and is quite another process from conduction and radiation. In many cases convection may greatly *aid* in cooling or warming bodies. But in all cases the body *actually receives or gives up its heat only by conduction or radiation*

from or to the body that has been moved near it. Thus, the water next to an automobile cylinder wall receives heat by conduction. The heated water circulates to the radiator where it gives up the heat by conduction to the metal walls, which in turn lose the heat by conduction to the air and by radiation. By convection the heated air is carried from the automobile.

103. General Statement of Thermal Conductivity. — Let heat be supplied at a steady rate to the point A of a body (Fig. 83). The heat will be conducted away from the point to the other portions of the surface of the body where it will be radiated or will pass into the surrounding medium by conduction. If the surrounding medium is kept at a constant temperature, finally a steady rate of flow of heat from the point A will be established. Through any two small areas B and C which are perpendicular to the direction of flow, energy will pass at a given rate. It is an experimental fact that the rate of flow of the heat between B and C is directly proportional to the temperature difference $d\theta$ existing between those surfaces, is directly proportional to their area A, and is inversely proportional to their distance apart ds. This may be expressed in the following equation

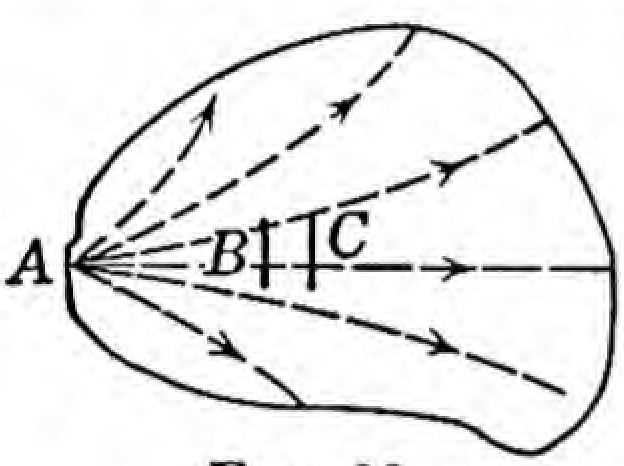


FIG. 83

$$\frac{dQ}{dt} = -k\frac{d\theta}{ds}A. \tag{156}$$

As s increases, the temperature decreases, hence $d\theta/ds$ is a negative quantity. A negative sign is used to make the right-hand member of the equation positive. The constant k is called the thermal conductivity of the medium. $d\theta/ds$ represents the space rate of change of temperature and is called the temperature gradient. The thermal conductivity is thus numerically equal to the quantity of heat transferred per second through a unit area at a place where the temperature gradient is unity.

104. Conduction of Heat along a Bar of Uniform Cross Section. — Consider a bar of length l with heat flowing at a certain uniform rate into one end and out at the other end. We shall consider that no heat flows through the side walls. Under these conditions dQ/dt is a constant for all cross sections A of the bar and its value may be obtained by measuring the finite quantity of heat Q which flows through the bar in a finite time t. Let the hotter end of the

bar have the temperature θ_1 and the colder end θ_2. Under the given conditions Eq. 156 becomes

$$d\theta = -\frac{Q}{kAt}\,ds.$$

In this equation each term of the expression Q/kAt is constant, so the integration is simple.

$$\int_{\theta_1}^{\theta_2} d\theta = -\frac{Q}{kAt}\int_0^l ds,$$

$$\theta_1 - \theta_2 = \frac{Q}{kAt}\,l,$$

and

$$k = \frac{Ql}{A(\theta_1 - \theta_2)t}. \qquad (157)$$

The above arrangement offers one of the simplest methods of determining the heat conductivity of materials. Heat is usually applied electrically at one end of the bar and various methods are used to avoid heat flow out through the sides or to reduce it to a negligible amount. When θ_1 and θ_2 reach steady values, the rate of heat flow must be constant and measurements are then made of the quantities needed in Eq. 157.

105. Conduction of Heat Radially through a Circular Cylinder. — The case of heat flow through a cylinder is of practical importance on account of the wide use of thick asbestos lagging on steam pipes as well as because it offers a precision method of measuring thermal conductivity.

Consider a hollow cylinder of the material whose thermal conductivity is to be measured. Consider that heat be supplied at a constant rate to the inside of the cylinder and let arrangements be made to keep the outside of the cylinder at a constant temperature. Soon a state of equilibrium will be reached when the inside and outside walls are maintained steadily at temperatures θ_1 and θ_2, respectively. The heat may be supplied at the center by an electrical heating unit made so as to supply heat evenly on all sides and of the same amount over each unit length of the cylinder. Let I be the current through the heater and E be the potential difference across its terminals. When equilibrium has been reached, the amount of heat flowing through the walls per second is obviously EI by Joule's Law (§ 147). Thus the constant dQ/dt is determined. Consider the cross section of the cylinder of

length l, inner radius r_1, and outer radius r_2 (Fig. 84). Consider a cylindrical shell of radius r and thickness dr which is within the cylinder. The surface area of this cylinder through which the heat flows is $2\pi rl$. Remembering that

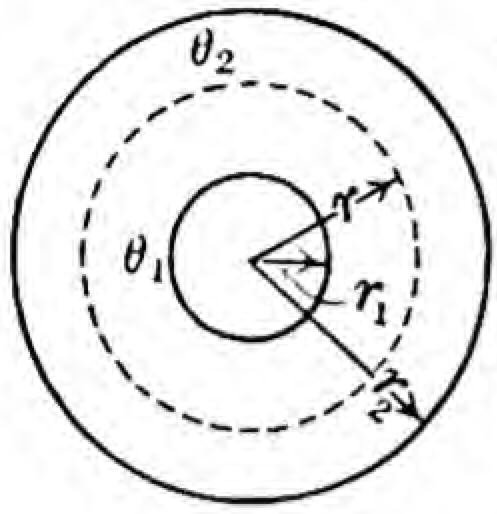


FIG. 84

$$\frac{dQ}{dt} = \frac{Q}{t} = EI \text{ is a constant,}$$

we obtain from Eq. 156,

$$\frac{Q}{t} = -k\frac{d\theta}{dr}2\pi rl.$$

Separating variables and integrating between the proper limits,

$$\int_{\theta_1}^{\theta_2} d\theta = \frac{-Q}{t\,2\pi kl}\int_{r_1}^{r_2}\frac{dr}{r},$$

$$\theta_2 - \theta_1 = -\frac{Q}{2\pi klt}\log_e\frac{r_2}{r_1}.$$

So
$$k = \frac{Q}{2\pi lt(\theta_1 - \theta_2)}\log_e\frac{r_2}{r_1}$$

$$= \frac{EI}{2\pi l(\theta_1 - \theta_2)}\log_e\frac{r_2}{r_1}. \tag{158}$$

The mechanical measurements of r_1, r_2, and l can be readily made and EI can be measured very accurately by potentiometer measurements (less accurately with voltmeter and ammeter). With sensitive thermocouples soldered to the inside and outside of the cylinder, $(\theta_1 - \theta_2)$ may be measured with good precision. Hence the errors in determining k by this method will not be due to the measurement of such simple quantities but will be due to the fact that the experimental arrangement cannot be kept as ideal as described above. The heating coil may not distribute the heat with perfect uniformity over the whole cylinder and only with special precaution can heat be kept from escaping in fairly large amounts from the ends of the cylinder.

106. The Determination of Thermal Conductivity in Other Regularly Shaped Containers. — Many substances cannot be made in cylindrical shape. For materials made in brick shape or flat slabs the best shape is probably a hollow cube. If the dimension of the faces is large compared with the thickness, then the heat flow at the corners can be neglected. The heat may be applied electrically and the temperature gradient through the wall meas-

ured by thermocouples placed on the inner and outer surfaces Eq. 157 may easily be applied to the case.

PROBLEMS

1. Show that for a cube of material, as described in the last paragraph, with edges l cm. long and the thickness of the walls d cm.,

$$k = \frac{Qd}{6\,l^2(\theta_1 - \theta_2)t}.$$

2. Show that for a spherical shell of outer radius r_2 and inner radius r_1 with heat applied electrically at the center,

$$k = \frac{EI}{4\,\pi(\theta_1 - \theta_2)}\left(\frac{1}{r_1} - \frac{1}{r_2}\right).$$

3. Show that in the metric system the units of k are calories per second, per ° C., per cm.; or joules per second, per ° C., per cm.

4. The conductivity of aluminum at 18° C. is 0.480 calories per second, per ° C., per cm. How much will its conductivity be when expressed in B.T.U. per hour, per square foot for a temperature gradient 1° F. per inch of thickness?

5. If k for concrete is from 6 to 9 in the English system of units as given in Problem 4, how much heat per hour would enter a refrigerator room 50 × 50 × 10 ft. with 6-inch concrete walls? The outside temperature 100° F., the inner temperature 25° F.

The Radiation Laws and the Measurement of High Temperatures

107. Newton's Law of Cooling. — There are certain quantitative relationships or laws between the energy radiated from hot bodies and the temperature of such bodies. These laws are the basis of high-temperature thermometry. Our entire knowledge of the temperature of incandescent bodies depends upon their validity. The earliest formulated law about heat losses is known as Newton's law of cooling. It states that for small temperature differences between a body and its surroundings the heat loss, per sq. cm. of surface, per second, is proportional to the temperature difference. In mathematical form

$$Q = RA(T - T_s)t, \tag{159}$$

or

$$\frac{dQ}{dt} = RA(T - T_s),$$

where Q is the heat given off, A the surface area, T the temperature of the body, T_s that of the surroundings, t the time, and R a constant depending upon the material. In this case, the heat Q is

partly due to radiation but mostly due to conduction and convection. The air next to the body is heated by conduction; this warm air is displaced by cooler air and thus air currents carry away the heat.

The rate at which Q increases may be computed from the heat capacity, C, of the body and the rate of decrease in its temperature (dT negative).

$$\frac{dQ}{dt} = - C\frac{dT}{dt}.$$

Then

$$C\frac{dT}{dt} = - RA(T - T_s),$$

$$\frac{dT}{T - T_s} = - \frac{RA}{C}dt.$$

Integrating,

$$\log_e(T - T_s) = - \frac{RA}{C}t + b. \tag{160}$$

Newton's law may therefore be tested by observing the temperature of the body at frequent intervals of time and plotting the values of log $(T - T_s)$ against the values of the time t. Such experiments give a linear relation if $T - T_s$ is not more than 20° or 30° C. The slope of this curve is seen to be RA/C, so if the heat capacity of the body and its surface area are known, then R, the heat lost from each cm^2., per second, per degree of temperature difference between it and the surrounding air, is easily determined. The constant R is important in making corrections for heat losses occurring during heat experiments.

108. Kirchhoff's Law. — *The ratio between the emissivity and the absorbing power is the same for all bodies and equal to the emissivity of a perfect absorber or "black-body" at the temperature in question.* The *emissivity* of a body is the energy radiated per sq. cm. per second from the body. The *absorbing power* is the ratio of the energy absorbed to that incident. For a perfect absorber this ratio is unity. Most surfaces do not absorb all the energy incident, part being reflected and part transmitted, the amount transmitted depending upon the transparency of the material. Surfaces which are black, such as carbon, lampblack, etc., are very nearly perfect absorbers, hence the name "black-body" for a perfect absorber. For all bodies that are not "black-bodies," the emissivity and absorbing power depend on the wave-length. Thus, a yellow pigment will absorb more energy from all the other regions of the visible spectrum than it does of yellow, and so yellow predominates

in the light which is not absorbed in the reflecting process. Likewise, as will be taken up in § 112, it is found by examining the spectrum of the light radiated from a hot substance, that there are different amounts of energy radiated in different parts of the spectrum. We may write Kirchhoff's law as

$$\frac{e_\lambda}{a_\lambda} = E_\lambda, \tag{161}$$

where e_λ is the emissivity of the surface for wave-length λ, a_λ the absorbing power for the same wave-length, E_λ a constant for a given temperature. E_λ is the emissivity of a "black-body," since when a_λ is unity the body is a perfect absorber and e_λ, the emissivity, then equals E_λ. We may write Eq. 161 as $e_\lambda/E_\lambda = a_\lambda$. The ratio e_λ/E_λ is called the relative emissivity. It is equal to the absorbing power and is found to be very nearly independent of the temperature.

109. Uniform Temperature Enclosures. — It may be shown as a consequence of Kirchhoff's law that the radiation from an enclosure whose walls have the same temperature at every point, is "full" or "black-body" radiation. The energy incident from without upon a small opening into an enclosure, will necessarily all pass in. The absorbing power of the enclosure, therefore, is unity and for temperature equilibrium the emission from the opening must, by Eq. 161, be that of a "black-body." It will be instructive to subject the radiation in such an enclosure to the following tests. Let us imagine as the most extreme case that the interior surface is highly reflecting. Such a surface in the open would radiate only a small fraction of the amount of energy that would come from a black surface at the same temperature. Now let us place a piece of carbon in the enclosure (Fig. 85). When the carbon has come to temperature equilibrium with its surroundings, it will be in a state where it is absorbing all the energy incident upon it and radiating an equal amount. In other words, the stream of radiant energy to the carbon from the surroundings will be exactly equal to the energy stream away from the carbon. The carbon becomes, therefore, indistinguishable. Likewise a piece

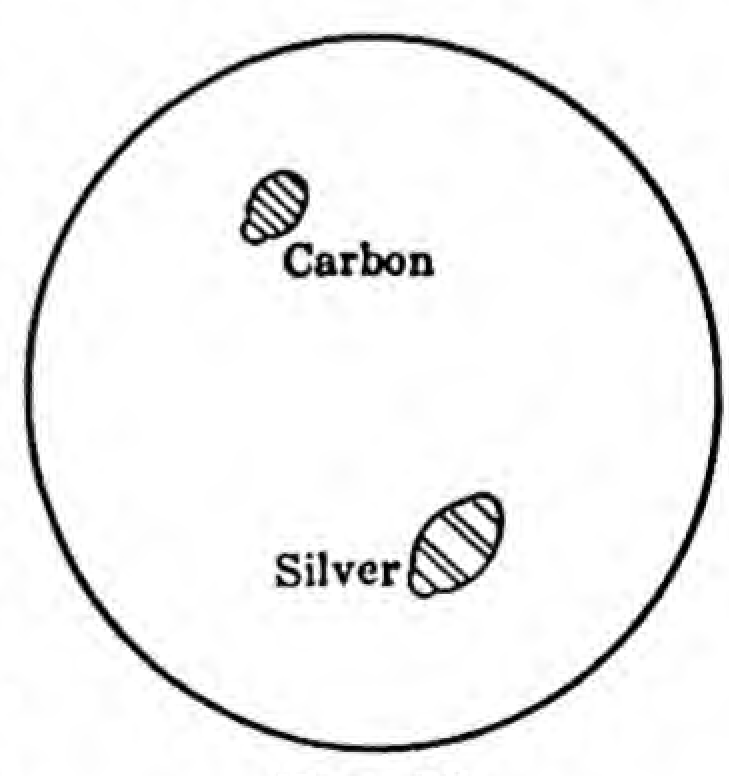


FIG. 85

of shiny silver placed within the enclosure becomes indistinguishable as soon as it attains the surrounding temperature. If these facts were not so, temperature changes would be occurring within the enclosure and this is contrary to our hypothesis. It is a familiar fact that in a furnace chamber in which temperature equilibrium is established, no outlines or objects can be distinguished. The same is true of the chinks between the glowing coals of a grate fire. The radiation, therefore, which escapes through a small opening is *independent of the material or surfaces within and is characteristic of the temperature only.* This fact is of great importance since the general radiation laws can be applied only to radiation which is a function of the temperature alone. A uniform temperature enclosure enables us to secure such radiation.

110. Stefan's Law. — *The radiation from a "black-body" is proportional to the fourth power of the absolute temperature,* that is

$$e = \sigma T^4, \tag{162}$$

where σ is a constant and e is the emissivity.

If the temperature of the body is T_1 and that of the surroundings T_2, the body is receiving energy at the rate $e_1 = \sigma T_1^4$. The net energy radiated in such a case is given by the expression

$$E = \sigma(T_1^4 - T_2^4). \tag{163}$$

This law is the basis of several temperature measuring instruments used in high temperature work. Among these may be mentioned the bolometer and the Fery pyrometer. The bolometer consists of a blackened strip of platinum used as one arm of a Wheatstone bridge. The radiation received by the blackened strip raises its temperature, the resulting resistance change is measured by the bridge, and the incident energy is computed. In the Fery pyrometer the radiation from a furnace opening is focussed by means of a mirror upon a sensitive thermo-junction. The resulting thermal e.m.f., being a measure of energy received, may be interpreted in terms of furnace temperature.

That Newton's law of cooling may be applied as an approximation to radiant energy alone may be shown as follows:

$$E = \sigma(T_1 - T_2)(T_1^3 + T_1^2T_2 + T_1T_2^2 + T_2^3).$$

Remembering that T represents Absolute temperatures, it will be seen that T is quite large even for bodies not far above room temperatures. For cases where T_1 and T_2 are both large and yet not very different, each expression in the last parenthesis is approxi-

mately T_1^3. So $E = 4\,T_1^3\sigma(T_1 - T_2)$, which is Newton's law. We may then see that if Newton's law is experimentally true for radiation, conduction, and convection combined, then if it is proved that the law is approximately true for radiant energy alone, it follows that the heat loss due to conduction and convection combined must also be approximately proportional to the temperature difference under the same limitations.

111. Wien's Displacement Law. — From the time of the ancients the color of a hot body has been used as a rough basis of temperature estimation. In certain industries today a statement of color, such as faint red, cherry, orange, white, etc., is a sufficient estimate of temperature. There is a physical law behind this which was first formulated by Wien; namely, for black-body radiation the wave-length of maximum energy (see the energy distribution curves, Fig. 86) shifts toward the shorter wave-lengths as temperature rises, and the product of wave-length of maximum energy and temperature is constant; that is,

$$\lambda_m T = \text{constant.} \quad (164)$$

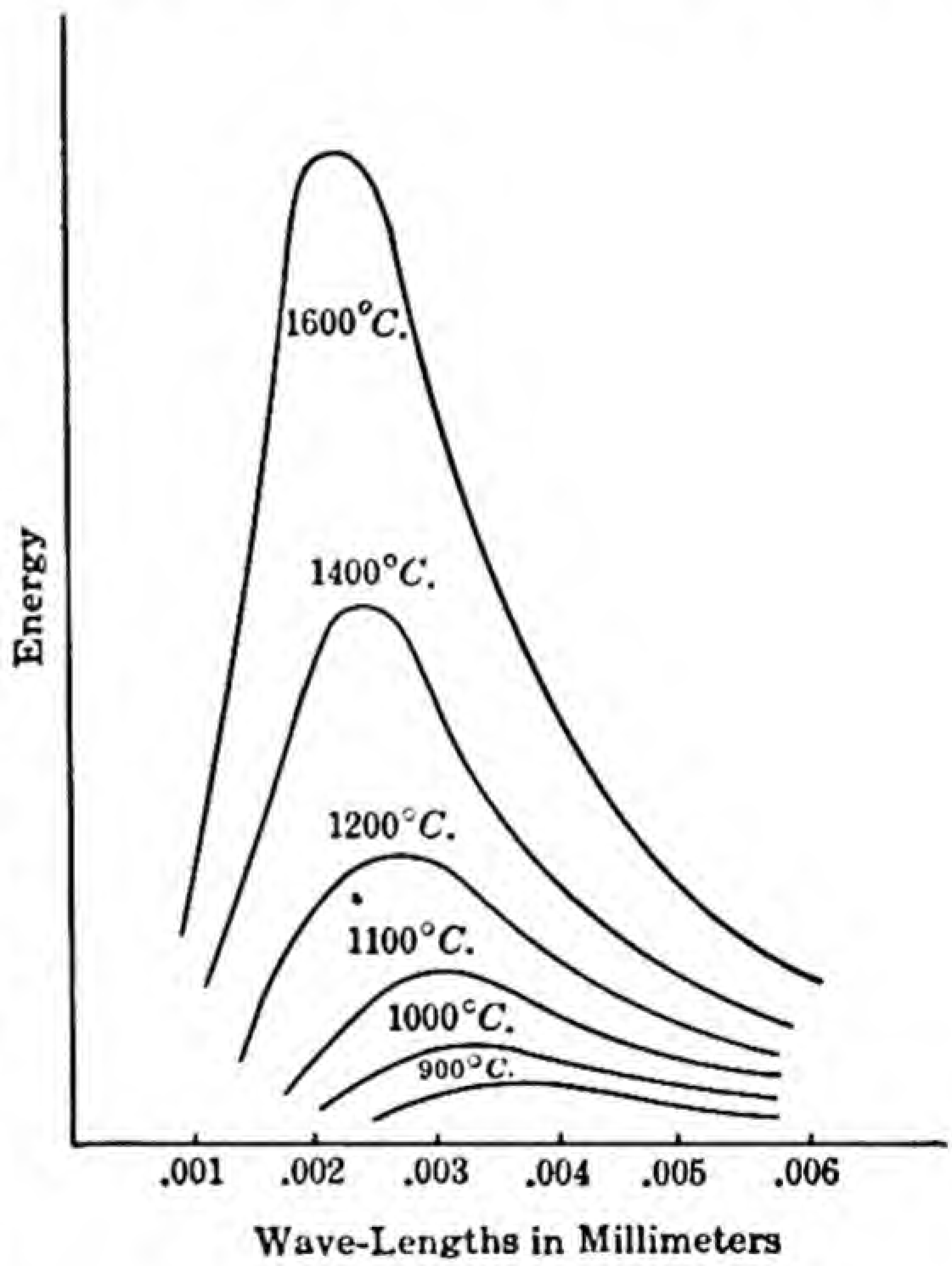


FIG. 86

The law does not give us a precise method of measuring temperature because of the practical difficulty in determining the wave-length of maximum energy, — especially at lower temperatures where the tops of the curves are quite flat.

112. Wien's Energy Distribution Law. Planck's Law. — This law first formulated by Wien and derived on the basis of thermodynamic reasoning gives the relation between the energy of any wave-length, that is, the energy comprised within any narrow band of the spectrum, and the temperature; namely,

$$E = c_1\lambda^{-5}e^{-\frac{c_2}{\lambda T}}. \quad (165)$$

λ is the wave-length chosen, c_1 and c_2 are constants, e is the base of the Napierian logarithms.

This law has been found to hold accurately only for the short wave-lengths of the visible spectrum. A modification proposed more recently by Planck has been found to hold accurately over the whole spectrum from the extreme infra-red to the ultra-violet. (See § 251.) Written according to Planck the law is

$$E = c_1\lambda^{-5}(e^{\frac{c_2}{\lambda T}} - 1)^{-1}. \tag{166}$$

For short wave-length this law reduces to Eq. 165.

Eq. 165 may be written for convenience in the form

$$\log_{10} E = \frac{K_1}{T} + K_2 \cdots \tag{167}$$

K_1 and K_2 are constants and T is temperature on the absolute Centigrade scale. This is the fundamental law of Optical Pyrometry.

113. Optical Pyrometry. — There are a number of instruments which make use of the above law in the measurement of high temperatures by optical methods. All make use directly or indirectly of some photometric method of comparing the intensity of the radiation from a hot body with that from some source whose temperature is known. If the ratio of intensities is known and the temperature of the comparison source, it is obvious from Eq. 167 that the unknown temperature may be determined. Since it is only necessary to compare intensities and not match them it is possible to determine temperatures many times hotter than that of the controllable source. In this way the temperature of the carbon arc has been determined and that of the sun and the fixed stars.

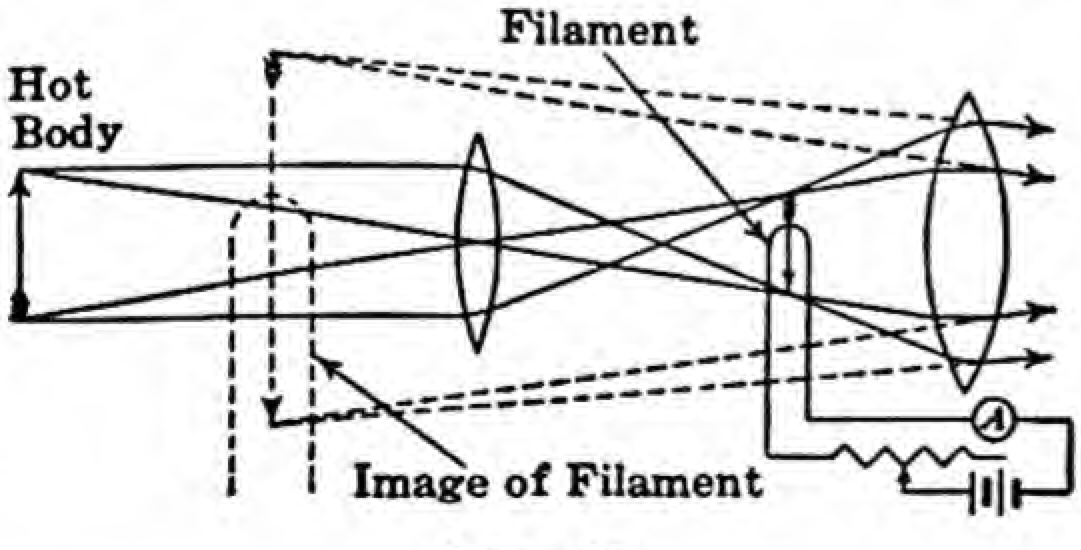


FIG. 87

In the Morse (Leeds and Northrup) pyrometer, one of the best known of these instruments, the temperature of a furnace chamber or hot body is obtained by matching the intensity of the radiation from the furnace or hot body with that of an incandescent lamp filament placed in the line of sight (see Fig. 87). The current through the lamp is adjusted until the filament disappears against

the luminous background. The intensities are then the same and if both bodies are full radiators, the temperatures are the same. The lamp filament is usually calibrated by sighting on a "black-body" whose temperatures are measured by some other means (such as resistance thermometer or thermo-junction) and noting the current that corresponds to various temperatures. Temperatures beyond the range of the calibration may be obtained by interposing a rotating sectored disk or an absorbing glass between the lamp and the body whose temperature is sought. In this way the intensity to be measured is cut down a definite known amount. If the apparent intensity, *i.e.* with the sectored disk or absorbing screen interposed, is E_a, while the actual intensity is E, we may write from Eq. 165

$$\frac{E}{E_a} = \frac{c_1\lambda^{-5}e^{-\frac{c_2}{\lambda T}}}{c_1\lambda^{-5}e^{-\frac{c_2}{\lambda T_a}}},$$

or

$$\frac{E}{E_a} = e^{-\frac{c_2}{\lambda}\left(\frac{1}{T}-\frac{1}{T_a}\right)},$$

where T_a, the apparent temperature, is measured and T, the actual temperature, is the only unknown, the ratio E/E_a being known. If the pyrometer is calibrated against a "black-body" it only gives true temperatures when the body sighted on is "black." If sighted upon a body radiating freely in the open, the pyrometer gives what is called the "black-body temperature"; that is, the temperature at which a "black-body" would radiate that type of radiation. The actual temperature in such a case is always higher than that given by the pyrometer and may actually be 300° or more above the pyrometer indication. If the emissivity of the surface is known, true temperature may be computed.

In the Wanner pyrometer one half of a photometric field is illuminated by light of a given wave-length from the hot body whose temperature is sought, the other half by light of the same wave-length from a comparison lamp. The two halves are adjusted to equal intensity by means of a polarizing device, the angle through which the Nicol prism is turned being calibrated in terms of temperature. This instrument is a direct application of Wien's law. Various other optical and radiation methods are in use.

CHAPTER IV

ELECTRICITY AND MAGNETISM

Electrostatics

114. Coulomb's Law. — Coulomb found by experiment that the force between electric charges varied directly as the quantity of each charge and inversely as the square of the distance between them. This law may be expressed thus:

$$F = \frac{QQ'}{kr^2}, \tag{168}$$

where F is the force, Q and Q' the two quantities of electricity, respectively, and r the distance between the two charges. The proportionality constant k depends upon the units selected for F, Q, and r and upon the medium in which the charges are placed.

Dielectric Constants

Vacuum	1.00000
Air	1.00059 at 0° C., 1 atmosphere
Air	1.00189 at − 186° C.
Hydrogen	1.00026 at 0° C., 1 atmosphere
N_2O	1.00113 at 0° C., 1 atmosphere
N_2O	1.010 at 15° C., 10 atmospheres
N_2O	1.025 at 15° C., 20 atmospheres
Sulphur	4.05 freshly crystallized
Sulphur	3.80 several months old
Glass (crown)	6.6
Mica	6.0
Hard Rubber	2.5
Pure Water	80.

The electrostatic unit of quantity of electricity is defined to be such that two such equal and like charges placed in a vacuum one centimeter apart repel each other with the force of one dyne. Because of this manner of defining the unit quantity of electricity, $k = 1$ for a vacuum. When the charges are surrounded by any matter, $k > 1$.

The proportionality constant k is called the *dielectric constant* of the medium. The value of k depends on the temperature, pressure, crystalline state and, in the case of alternating fields, upon the frequency of alternation.

115. The Electrostatic Field. — An electric field is said to exist at a given point if a charge placed at that point is acted upon by a force. Theoretically there is a field around any charge which extends out to infinity, but when r gets so large that the force becomes too small to detect, the field may be considered to be non-existent.

It is an experimental fact that the force on a charge placed at a given point in an electric field varies directly as the amount of the charge. It is arbitrarily agreed to measure the strength or intensity of the field by means of the force exerted on a given charge placed in the field, — thus one electric field is said to be twice as intense as another, if a given charge is acted upon with twice as much force when placed in the one field as when placed in the other field. We may, therefore, say that the force on a charge in an electric field varies directly as the strength of the field, E, and directly as the amount of the charge, Q. $F = bQE$. We may make the proportionality constant b equal to unity by properly defining the unit value of E.

An electric field at a given point is said to have an intensity of one electrostatic unit (1 e.s.u.) of field strength when an electrostatic unit of quantity placed at that point is acted upon by the force of one dyne. The direction of the field is arbitrarily chosen to be the direction in which a positive charge is urged. Since the intensity of an electric field has both magnitude and direction, it is a vector quantity.

From this definition it follows that a charge Q when placed in a field of intensity E is acted upon by a force

$$F = QE \text{ dynes.} \tag{169}$$

This equation is the defining equation for E. Hence *the intensity of an electric field (in e.s.u.) at a point is equal to the force (in dynes) which would act on a unit charge (1 e.s.u.) placed at that point.*

As a particular case, we see from Eq. 168 that the intensity of the field at a distance r from a point charge Q is

$$E = \frac{Q}{kr^2} \text{ e.s.u. (or dynes per unit charge).}$$

116. The Cavendish Proof of the Inverse Square Law. — The experiments of Coulomb showing that the force of attraction between electrostatic charges varies inversely as the square of the distance between the charges were not very accurate. In fact, direct measurements of the forces cannot prove the law accurately for the following reasons. When charges are alike, they repel each other, so if they are placed on small conducting spheres, the charges tend to concentrate on the back sides of the spheres as the spheres approach each other. Then although the forces acting on the spheres may be measured with high accuracy, the distance between the charges is indeterminate. The effective distance is larger than the distances between the centers of the spheres, but the exact center where the total charge might be considered concentrated is not determinable. The following very accurate, though indirect, proof was suggested first by Cavendish.

The force between the charges is some function of the distance between them. The Cavendish proof assumes that the function is of the form $1/r^n$. To prove that $n = 2$, use is made of the fact that at every point inside a closed conductor there is no electric field,—*i.e.*, there is no force on a charge placed at any point inside.

First let us consider some geometrical properties of a sphere. The solid angle subtended at the center of a sphere of one cm. radius by one sq. cm. of surface is called a unit solid angle, or one steradian. Since the surface of a sphere varies as the square of the radius, the area subtended by one steradian on the surface of a sphere of radius r is r^2 sq. cm. So any spherical surface area divided by r^2 gives the number of units of solid angle subtended at the center.

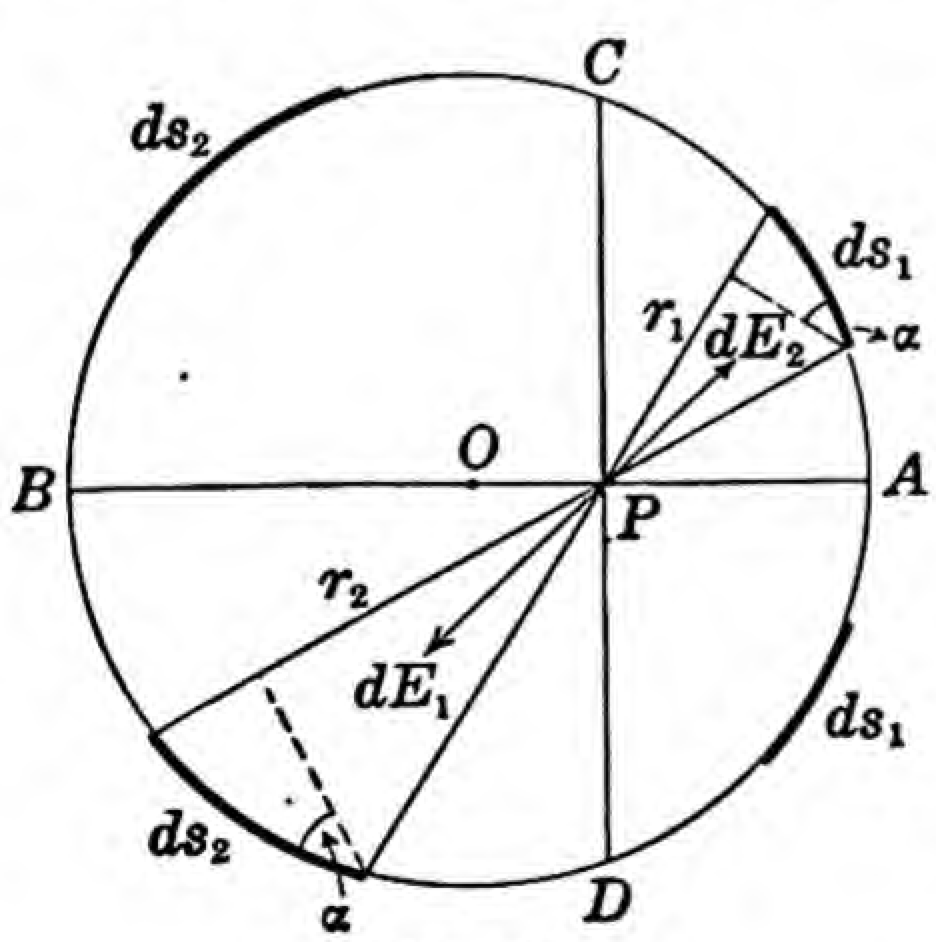


FIG. 88

Let O (Fig. 88) be the center of a sphere and P be any point inside the sphere. Let a line through P generate a cone of infinitesimal size, which cuts out on the surface of the sphere the areas ds_1 and ds_2. Let r_1 and r_2 be the distances of these surfaces from P, respectively. The areas ds_1 and ds_2 make equal angles, α, with the planes perpendicular to the axis of the cone. Since the

areas ds_1 and ds_2 are infinitesimals, the projected areas $ds_1 \cos\alpha$ and $ds_2 \cos\alpha$ may be considered the same as surfaces of spheres whose radii are r_1 and r_2, respectively. The solid angles at P are equal to each other, so

$$\frac{ds_1 \cos\alpha}{r_1^2} = \frac{ds_2 \cos\alpha}{r_2^2},$$

or

$$\frac{ds_1}{r_1^2} = \frac{ds_2}{r_2^2}. \tag{170}$$

This relationship will be used shortly.

Consider the above sphere to be a charged conductor. The mutual repulsion of the individual charges causes a uniform distribution of charge over the surface. Let q be the charge per square centimeter. The charge on ds_1 is qds_1 and that on ds_2 is qds_2. The strengths of the electric fields at P due to these charges are

$$dE_1 = \frac{q\,ds_1}{r_1^n}, \quad \text{and} \quad dE_2 = \frac{q\,ds_2}{r_2^n}. \tag{171}$$

These fields, for the sphere charged positively, are indicated in the figure by the arrows. Comparing Eqs. 170 and 171 we see that if $n = 2$, $dE_1 = -\,dE_2$ and likewise for all the other elements, so the resultant electric field at P will be zero.

Since it might possibly be true that the electric field would be zero no matter what the value of n might be, we shall next investigate the value of E when $n \neq 2$. Pass a plane CD through the sphere perpendicular to the line through O and P. Let all the elements ds_2 be restricted to the surface lying on the left of the plane CD. Then the elements ds_1 will make up the remainder of the sphere, and r_2 will always be larger than r_1.

Let the resultant field at P due to the charge on ds_1 and ds_2 be dE.

$$dE = dE_1 - dE_2 = q\left(\frac{ds_1}{r_1^n} - \frac{ds_2}{r_2^n}\right).$$

Case I. $n > 2$. — It must be remembered that when $n = 2$,

$$\frac{ds_1}{r_1^n} = \frac{ds_2}{r_2^n} \quad \text{or} \quad dE_1 = dE_2.$$

Since r_2 is greater than r_1, then as n becomes greater than 2, the denominator r_2^n increases much more rapidly than r_1^n and the fraction ds_2/r_2^n becomes smaller than ds_1/r_1^n. Therefore the field dE_1 becomes larger than dE_2.

For each element ds_2 on the lower half of the sphere there may be selected an equal element located symmetrically on the upper half of the sphere. The charges on these two symmetrical elements will create fields at P having the same magnitudes and making equal angles with the line BA. The components of these fields perpendicular to BA will be directed oppositely and will cancel while the components parallel to BA (directed away from the center O) will add. Likewise the charges on the two symmetrically located elements ds_1 will produce fields whose components perpendicular to BA cancel and whose components parallel to AB (directed toward the center O) will add. Since $dE_1 > dE_2$, the resultant field at P due to these four elements will be directed from P toward O.

When the effect of all the charges on the surface of the sphere is considered, it is seen that the resultant field at P is directed toward the center of the sphere.

Case II. $n < 2$. — Because r_2 is greater than r_1, then as n becomes less than 2 the denominator r_2^n decreases much more rapidly than r_1^n and ds_2/r_2^n becomes larger than ds_1/r_1^n. Then the field dE_2 becomes larger than dE_1.

Considering the effect of all the charges on the sphere as in the above case, we see that the resultant field at P is directed toward A, away from the center of the sphere.

Thus did Cavendish prove that the field at any point inside a closed conducting sphere can be zero only if the inverse square law is perfectly true.

The experimental test for the absence of the electric field was made as follows:

A sphere A (Fig. 89) was mounted inside of a sphere B and insulated from it. The outer sphere was then given a large positive charge. With the lid L closed, the wire passing through it was made to touch the sphere A. The wire was withdrawn, the lid opened, and the inner sphere tested to see if any charge could be detected upon it. Cavendish predicted from the above argument that if $n > 2$ the field would force some of the positive charge from B to A. If $n < 2$, A would be negatively charged and if $n = 2$, A would be uncharged. Cavendish could detect no charge on A. Later Maxwell carefully performed the experiment with more delicate instruments and could

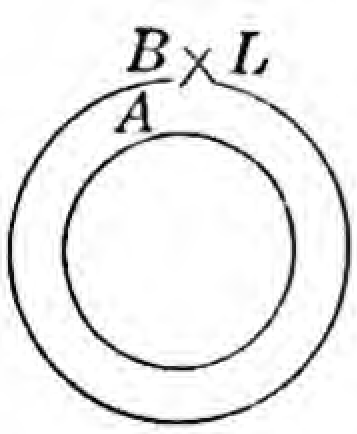


FIG. 89

detect no charge on A. Knowing the smallest charge his instruments would detect, Maxwell deduced that n was equal to two within one part in over 20,000.

117. Electric Induction. — It is customary to represent an electric field graphically by a set of lines which originate on positive charges and terminate on negative charges. These curved lines are drawn so that at any point a tangent to the curve gives the direction of the electric field at that point. The question arises: How many of these lines shall be drawn, in order to represent the field? Although an infinite number are possible, it has been decided to limit the density of the lines so as to indicate quantitatively the strength of the electric field. Thus it has been agreed that E lines will be drawn through a unit area at right angles to the direction of the field in order to represent a field of strength E. Thus in Fig. 90a at A, distant r from the charge Q, the strength of the electric field is $E_0 = Q/r^2$ dynes per unit charge and Q/r^2 lines will be drawn through the square centimeter shown in order to picture the magnitude of the electric force which would act on a unit charge placed at A.

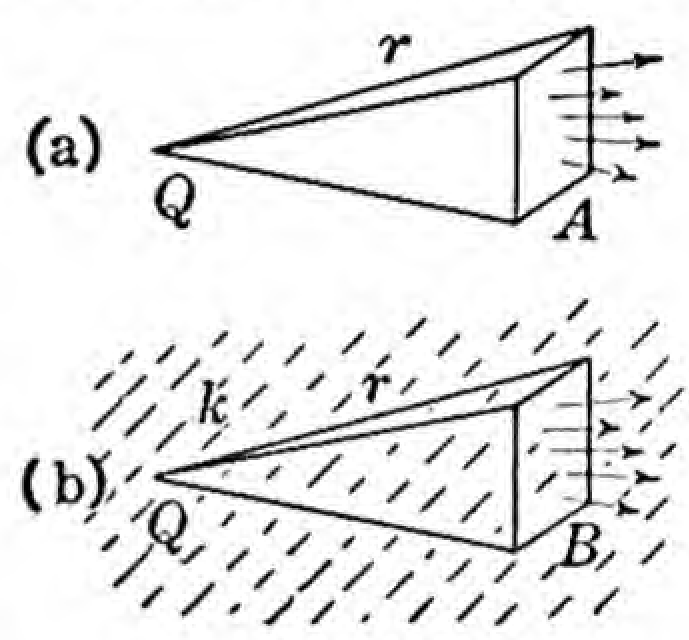


FIG. 90

Now suppose the whole space around the charge Q is filled with a medium of dielectric constant k, Fig. 90b. By Coulomb's law the field at B, distant r from the charge, is $E = Q/kr^2$. Now two choices are available for the picturing of the field by lines: either we may agree to draw only Q/kr^2 lines through each square centimeter at B (making the density of lines always equal to the field strength) or we may agree to draw Q/r^2 lines through each square centimeter (leaving the density of lines the same as in case (a)). The latter choice has been agreed upon. That choice was made because it results in having the same total number of lines radiate from a given charge regardless of the medium in which it is placed. In case (b) it is seen that k times as many lines are drawn through each square centimeter as there are dynes of force acting on a unit charge. Let N be the number of lines drawn in the medium through each square centimeter at right angles to the direction of the field. We thus have the formula,

$$N = kE. \tag{172}$$

N is called the induction or the flux density. The lines representing E are called lines of force. Those representing N are called lines of induction. It should be remembered that *both N and E represent the number of lines passing at right angles through one square centimeter of surface.*

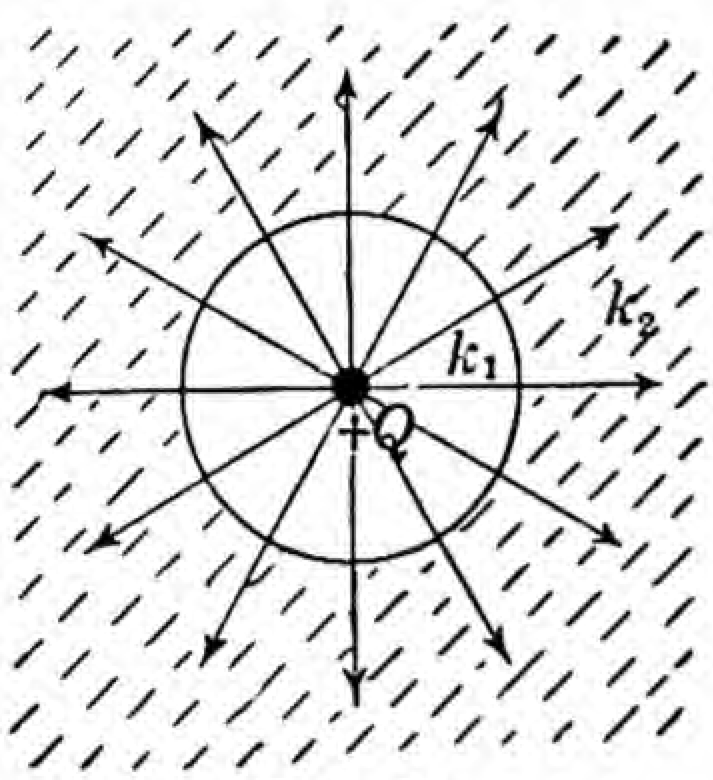


FIG. 91

By adopting the above convention of representing N, we are enabled in the case illustrated in Fig. 91 to have the number of lines remain constant when they pass from one medium to another. The lines may be bent, if they do not enter perpendicularly the boundary between the two media, but they never increase or decrease in number. Had the first of the above conventions been adopted, the number of lines would change and the lines would have to be discontinuous at the boundary.

118. Gauss's Theorem. — *The total number of lines of induction, or the total flux, across any closed surface is* 4π *times the total charge enclosed within the surface.*

Let a charge $+Q$ be located at O, inside a closed surface, Fig. 92. The strength of the field at any point on an element ds of the surface is

$$E = \frac{Q}{kr^2}$$

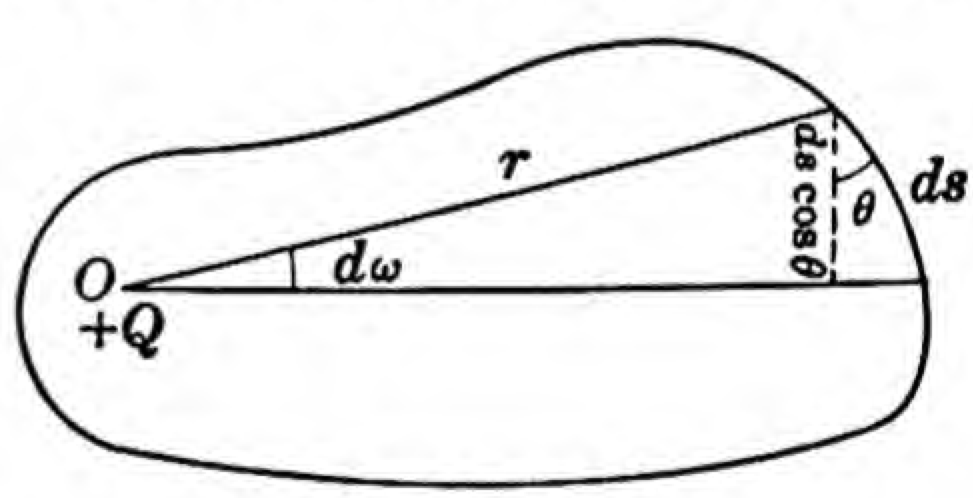


FIG. 92

and is directed outward along the radius vector r. The projected area $ds \cos \theta$ is at right angles to the direction of the field and we must represent the induction by kE lines through each square centimeter of the area. The *total normal induction* across the projected area is

$$k \frac{Q}{kr^2} ds \cos \theta \text{ lines.} \tag{173}$$

This same number of lines also passes through ds but not normally. The solid angle subtended at O by ds is

$$d\omega = \frac{ds \cos \theta}{r^2},$$

so, from Eq. 173, the *total induction, or flux,* $d\Phi$, across ds is $d\Phi = Q\,d\omega$. Then, integrating over the whole of the closed surface,

$$\Phi = \int d\Phi = \int_0^{4\pi} Q\,d\omega = 4\,\pi\,Q, \tag{174}$$

since there are $4\,\pi$ steradians about a point. It is to be noticed that Φ is the total number of lines regardless of their directions with respect to the surface.

The same equation will apply to any and all charges within the closed surface, so

$$\Phi = 4\,\pi\Sigma Q. \tag{175}$$

If some of the charges are negative, their lines of force will enter the surface and they must be subtracted from those leaving. So in Eq. 175 the algebraic summation must be taken.

This proof shows that because of the way of defining the induction in Eq. 172 every unit charge must be pictured with $4\,\pi$ lines emerging from it (or entering) regardless of the medium surrounding it.

Gauss's theorem is often stated in a form which might at first seem to be entirely different. Consider Eq. 173 in the following form:

$$k\left(\frac{Q}{kr^2}\cos\theta\right)ds = kE\cos\theta\,ds.$$

The term kE is the flux density or the induction and $kE\cos\theta$ is the component of the induction which is perpendicular to the surface. When this is multiplied by the element of area, ds, the expression is seen to represent the normal induction over an infinitesimal area of the surface. The integration of that expression gave $4\,\pi Q$. So we may state the theorem in the following form: The surface integral of the normal component of the induction over a closed surface is equal to $4\,\pi$ times the number of unit charges residing within the surface.

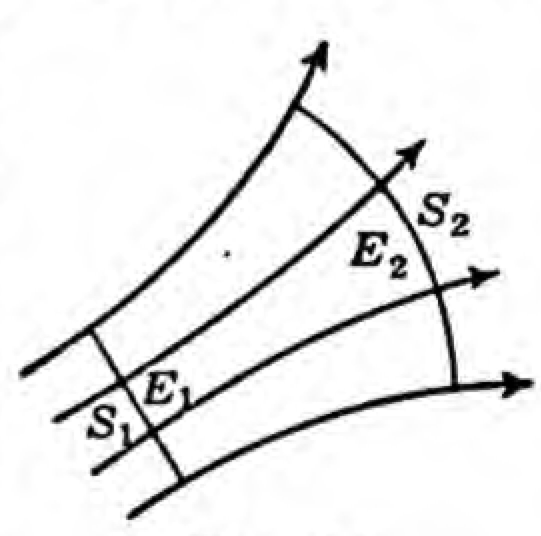


Fig. 93

In § 117 we agreed to make the density of lines of force equal to the numerical value of E. Now let us investigate the consequences of this convention. Consider an area S_1 (Fig. 93) which is perpendicular to the set of uniformly spaced lines of force of density equal to the field E_1. Now follow along these lines of force until they subtend an area S_2, all parts of which are at right angles to the

lines of force. The density of lines through S_2 is less than through S_1. The question arises: Is the density of lines of force through S_2 equal to the intensity of the field E_2? The faces S_1 and S_2 and the lines of force passing through their peripheries constitute a closed surface. We know from Gauss's theorem that since there are no charges within the surface, the total flux passing through the surface is zero. Therefore the flux passing in through S_1 equals the flux passing out through S_2. Therefore $\Phi_1 = \Phi_2$, or $N_1S_1 = N_2S_2$. Applying Eq. 172 we get $kE_1S_1 = kE_2S_2$ or $E_1S_1 = E_2S_2$. The answer to the question proposed above lies in these equations. From the latter equation we see that the values of E at the two surfaces are inversely proportional to the areas of the surfaces. Since the same lines of force pass through both S_1 and S_2, the density of the lines must likewise be inversely proportional to the areas S_1 and S_2. Therefore the values of E and the density of lines of force are proportional to each other. Having been made equal (by agreement) at the surface S_1, they must remain equal everywhere. Therefore when a set of lines are drawn so that their density equals the field strength at any one point, then their density represents the field strength at all points along their whole length.

119. Electric Potential and Potential Difference. — If an electric charge is moved into an electric field from some point outside the field (infinity), work will have to be done against the electric forces brought into play. *The work done on a unit positive charge in bringing it from infinity to any given point in a field is called the electric potential at that point.* It is the potential energy which a unit positive charge would possess at that place.

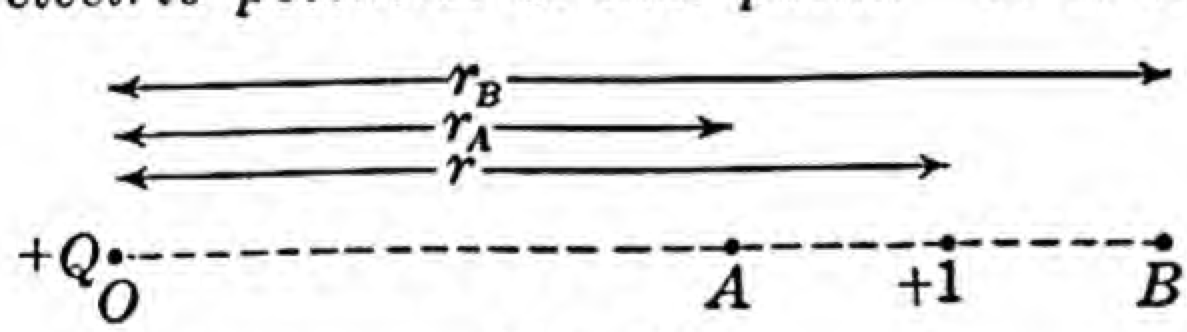


FIG. 94

In the case of an isolated charge concentrated at a point, the surrounding field is radial and the potential at any point is easily computed. If the unit positive charge is brought in from infinity along a given radius (Fig. 94) the force is always parallel to the displacement and the work $dw = E(-dr)$. If r represents the distance measured outward from O, then dr will be negative when the unit charge is moved inward toward O. The vector E is directed away from O, therefore $E\,dr$ is essentially negative for motion toward O. In order to have the work done on the positive

unit charge a positive quantity, the negative sign must be used. Calling V_r the potential at a point distant r from a charge Q,

$$V_r = -\int_{\infty}^{r} E\,dr = -\int_{\infty}^{r} \frac{Q}{kr^2}\,dr = \frac{1}{k}\frac{Q}{r}. \tag{176}$$

In most practical cases we are interested in knowing only the difference in the potentials between two points. *Two points are said to have one electrostatic unit (one e.s.u.) of potential difference when one erg of work is done in transferring one electrostatic unit of quantity of positive electricity from the point of lower to the point of higher potential.* The difference in the potentials at points A and B, distant r_A and r_B respectively from the charge $+ Q$, is

$$V_A - V_B = \int_{r_B}^{r_A} -\frac{Q}{kr^2}\,dr = \frac{1}{k}\left(\frac{Q}{r_A} - \frac{Q}{r_B}\right). \tag{177}$$

When Q is in e.s.u. and r_A and r_B in cm., then $V_A - V_B$ is in e.s.u.

If a conducting body were suddenly placed near the charge $+ Q$ so that part of it was at A and part at B, instantaneously the potential at A would be higher than at B. But two points of a conductor cannot remain at different potentials unless a steady current flows between the points, which demands a continuous supply of energy. So electrons will flow from the points of the body at lower potential to the points at higher potential until all parts of the body have the same potential. Thus the side next to the positive charge will be negatively charged and the opposite side positively charged.

If the path along which the unit positive charge is carried is not everywhere parallel to the electric field, then just as in §§ 2 and 15 we must find the potential difference between two points M and N as follows:

$$V = V_M - V_N = -\int_{N}^{M} E\cos\theta\,dr. \tag{178}$$

Wherever practical, the integration is carried out along a line of force so that $\cos\theta$ is everywhere unity.

A very useful relation may be obtained by noting that we may write $dw = dV = - E\,dr$ in the form:

$$E = -\frac{dV}{dr}. \tag{179}$$

This relation states that the intensity of electric field in a given direction is numerically equal to the rate of change of the potential in that direction and the negative sign indicates that the field

has the direction in which the potential decreases. The student should keep continually in mind the two ways of measuring the value of E; (1) by the force on a unit charge, (2) by the number of units of drop in potential per centimeter.

PROBLEMS

1. Show that the work done in carrying a charge from one point to another in an electric field is independent of the path or else the principle of conservation of energy would be violated.

2. In Fig. 91 let the sphere separating the two media have a radius of 5 cm. and let $q = +100$ e.s.u. Let $k_1 = 1$ and $k_2 = 4$. Compute the electric field strengths and the potentials at points just inside and just outside the boundary of the two media. Compute the potential difference between two points which are respectively 2 and 10 cm. from the center.

120. Capacity. — Consider two parallel conducting plates M and N of equal area separated by a distance s (Fig. 95). Let the plates be charged with equal and opposite charges, q per square centimeter, and let the potential difference across the plates be V. Over most of the space between the plates the lines of force joining opposite charges are nearly straight parallel lines, but approaching the edges one finds that the lines bulge outward more and more and the electric field becomes weaker. However, if the dimensions of the plates are large compared to the distance of the plates apart, the electric field will be found to be practically the same at all points between the plates. Then the lines of induction must be drawn parallel and $4\pi q$ lines must be drawn from each square centimeter of the plates. So Eq. 172, when combined with Gauss's theorem for this case, becomes

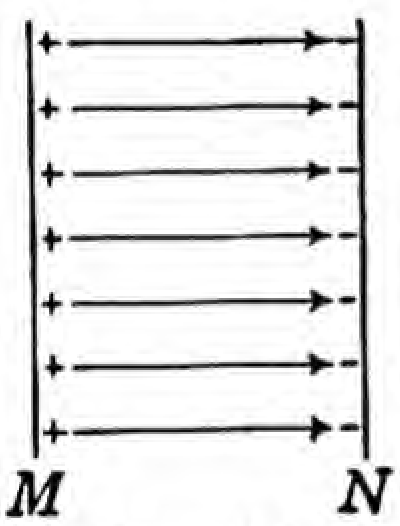


FIG. 95

$$N = 4\pi q = kE. \tag{180}$$

It is to be remembered that q in this equation is the charge *per unit area* of the plates.

We may now use Eqs. 178 and 180 to find the potential difference across the two plates in terms of their distance apart and the charge placed upon them. In the case of parallel plates E is constant and, from Eq. 180, equal to $4\pi q/k$. Choosing a line of force for the path of integration, $\cos\theta$ is unity at all times. Therefore, from Eq. 178,

$$V = \int_N^M E\,dr = E\int_N^M dr = Es = \frac{4\pi q}{k}s. \tag{181}$$

It is to be noted that from here on we deal mostly with potential difference and use the letter V without a subscript to stand for it. Also whenever just the magnitude of the potential difference is wanted the signs of E and dr are ignored and the order of the limits taken so that V is a positive quantity.

We now wish to consider the effect on the potential difference across the two plates (a) of moving the plates closer to each other and (b) of changing the dielectric constant of the medium between the plates. (See Fig. 96.)

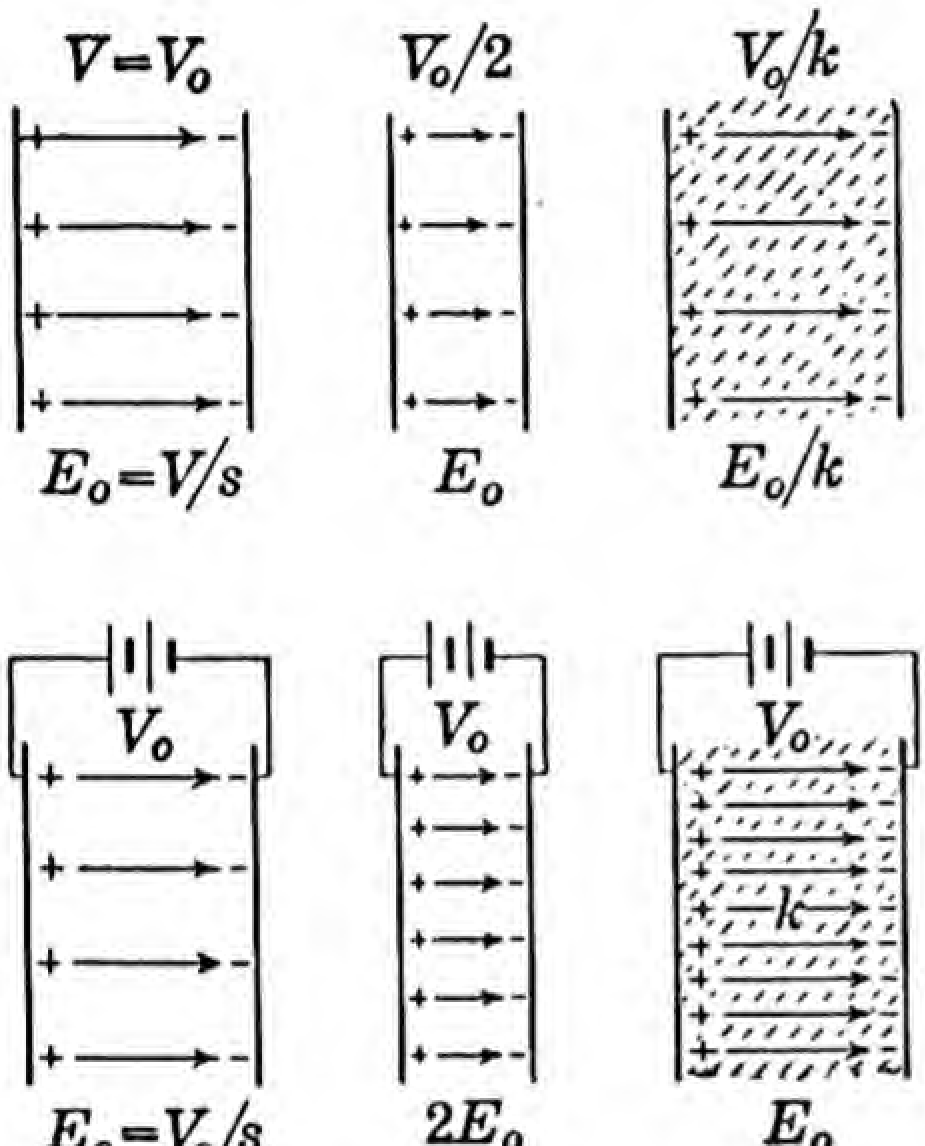


FIG. 96

Case I. Charge on the Plates Kept Constant. — (a) If the plates are insulated, then the charge q cannot change from its original value q_0 as the plates approach each other and hence N cannot change.

From Eq. 180 we see that if N is constant then E must also remain constant at its original value of E_0, *i.e.* the drop in potential per cm. is constant. Hence, if the distance between the plates is halved, then the total potential drop across the plates is halved. This is seen algebraically from Eq. 181: the value of E_0 is fixed, hence as s is decreased, the product of E_0s is decreased proportionately.

(b) Let the dielectric medium between the plates be changed from one whose dielectric constant is k_0 to one whose constant is k. The induction N is not changed because q is not changed, so $k_0E_0 = kE$. The intensity of the electric field becomes $E = k_0E_0/k$. If the original dielectric was air, then the field is reduced to nearly E_0/k. Therefore the potential difference is decreased from V_0 to nearly V_0/k.

Case II. Potential Difference of the Plates Kept Constant. — (a) An interesting and important set of facts is obtained when, instead of keeping q_0 constant, the value of V_0 is kept constant. This is managed by keeping the plates attached to a battery while they are moved or while the dielectric between them is changed. Eq. 181 shows that if $V = V_0$ is constant, then when s is decreased,

E must increase in direct proportion and Eq. 180 states that because E is increased the charge q must increase in the same proportion; *i.e.* more electricity flows from the battery onto the plates.

(*b*) Now let the distance between the plates be left unchanged but the dielectric changed from that of $k_0 = 1$ to one of value k. Since V_0 is kept constant, then $V_0/s = E_0$ is constant. Eq. 180 then states that q will be increased to k times its former value, and hence the number of lines of induction increases k times.

It is thus seen that the amount of charge on the plates of a condenser depends on the area of the plates, their distance apart, the material between them, and upon the potential difference of the plates.

When a condenser is charged or discharged, the motion of charge occurs through the outer metallic conductors and it is known that in metals only electrons can flow. The external battery causes all of the electrons in the circuit to be displaced slightly, producing an excess negative charge on one plate and a deficiency of electrons (a positive charge) on the other. So upon charge or discharge, the quantity of electricity which moves is only the quantity equal in magnitude to that on either single plate, and that is spoken of as the quantity of electricity or charge in the condenser. *The ratio of the quantity of electricity in the condenser to the potential difference across the plates is called the capacity of the condenser.* So

$$Q = CV, \tag{182}$$

and

$$C = \frac{Q}{V} = \frac{Q}{\int E\,dr}. \tag{183}$$

From this equation it is seen that if the value of E is known as a function of the distance r from one of the plates, then the denominator is integrable and the capacity is determinable. As will be shown below, for many condensers of symmetrical form, the value of E is obtained by a simple application of Gauss's theorem as stated in Eq. 180.

We have seen in analyzing the cases illustrated in Fig. 96 that when a medium of dielectric k fills the space between the plates, the quantity of electricity for a given potential difference V increases to k times the original value. So we see from Eq. 183 that if C_0 is the capacity of a condenser with nothing between the plates,

then when the space between the plates is filled with a medium of dielectric constant k the capacity of the condenser becomes $C = kC_0$.

PROBLEMS

1. Consider a spherical surface of radius 5 cm., with a + charge of 100 e.s.u. at its center. What is the value of the electric field, the normal induction, and potential at any point on the sphere? Also what is the total induction through the sphere? Answer the question first for the case where the medium is a vacuum and then for mica.

2. Two charges, +80 and +100 e.s.u., are placed 10 cm. apart in a vacuum. At a point halfway between them, what value have E and N? What force would be exerted on a charge of +10 e.s.u. placed there? Answer both questions for the case where the medium is water.

3. In the problem above replace the charge of +100 e.s.u. by one of −100 e.s.u. and solve.

4. In Problems 2 and 3, what is the potential at the midpoint and at a point 1 cm. closer to the 100 unit charge?

121. Capacity of a Parallel Plate Condenser. — Consider a condenser made of two parallel plates each of area A and separated by a distance s. Let Q be the charge on each plate. As discussed in the previous section, E is constant so Eq. 183 gives

$$C = \frac{Q}{Es} = \frac{Q}{\dfrac{4\,\pi Q}{kA}\,s} = \frac{kA}{4\,\pi s}\ \text{e.s.u.} \tag{184}$$

122. Capacity of a Spherical Condenser (Concentric Spheres). — Since the field between two concentric spheres (Fig. 97) is seen to consist of nonparallel lines, the field must be weaker at points more distant from the center. Therefore, before $\int E\,dr$ can be evaluated, the value of E at any point between the spheres must be found as a function of the distance r from the center. Let Q be the total charge on the condenser. By Gauss's theorem there are $4\,\pi Q$ lines of induction and because of symmetry we know they must be uniformly distributed. For a spherical surface of radius r, the $4\,\pi Q$ lines are distributed over $4\,\pi r^2$ cm²., so

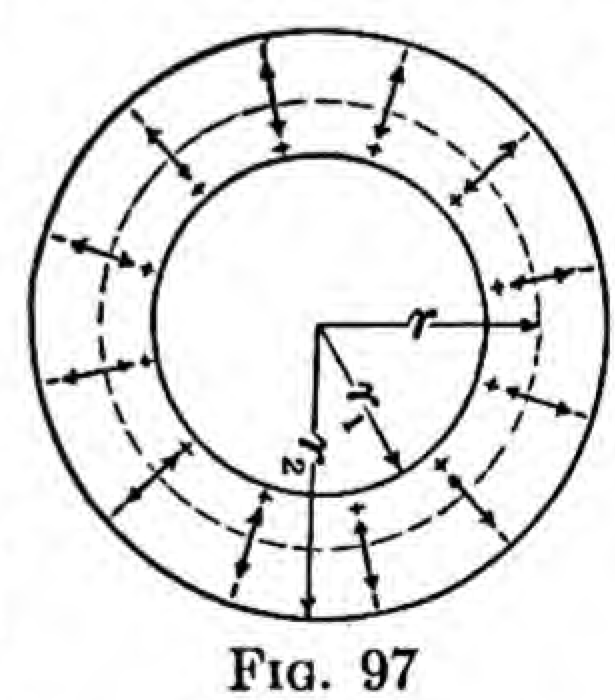


FIG. 97

$$kE = \frac{4\,\pi Q}{4\,\pi r^2} = \frac{Q}{r^2}\,\frac{\text{lines}}{\text{cm}^2.},$$

i.e. the field is the same as it would be if all the charge were concentrated at the center of the sphere. Eq. 183 then gives

$$C = \frac{Q}{\displaystyle\int_{r_1}^{r_2} \frac{Q}{kr^2}\,dr} = \frac{k}{\left.-\dfrac{1}{r}\right]_{r_1}^{r_2}} = \frac{k}{-\dfrac{1}{r_2} + \dfrac{1}{r_1}} = \frac{kr_1r_2}{r_2 - r_1} \text{ e.s.u.} \quad (185)$$

Let r_2 become large without limit. Then we have the case of an isolated charged sphere. For this case $C = kr_1$. So for $k = 1$ we have the capacity of an isolated sphere numerically equal to its radius in centimeters.

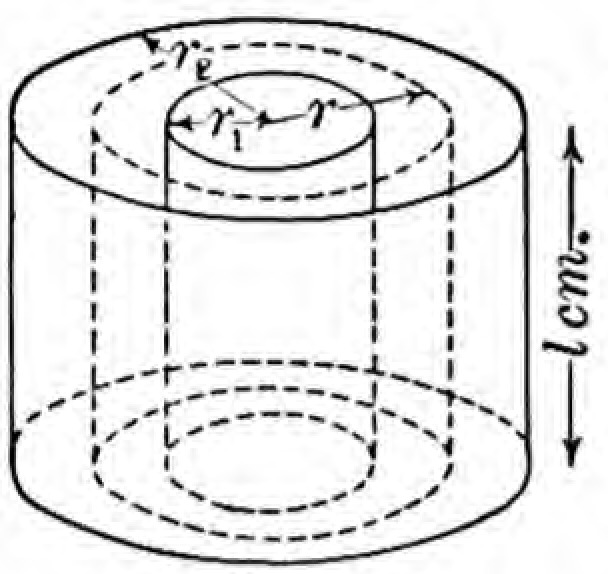


FIG. 98

123. Capacity of a Cylindrical Condenser. — Consider a condenser whose plates are two coaxial cylinders of length l. Let Q be the charge on each cylinder (Fig. 98). Consider a coaxial cylinder of radius r between the two plates of the condenser. Its surface area is $2\pi rl$. Through its surface must pass the $4\pi Q$ lines.

So
$$kE = \frac{4\pi Q}{2\pi rl} = \frac{2Q}{rl}\,\frac{\text{lines}}{\text{cm}^2}.$$

From Eq. 183,

$$C = \frac{Q}{\displaystyle\int_{r_1}^{r_2} \frac{2Q}{krl}\,dr} = \frac{kl}{\left. 2\log r\right]_{r_1}^{r_2}} = \frac{kl}{2\log\dfrac{r_2}{r_1}} \text{ e.s.u.} \quad (186)$$

124. Capacities in Parallel and in Series. — By way of review, formulae for these cases will be derived. The student should draw the necessary figures.

(*a*) *Parallel Arrangement.* — From our knowledge of electricity

we know that $\quad V = V_1 = V_2 = V_3 = \cdots$

and $\quad Q = Q_1 + Q_2 + Q_3 + \cdots$

By definition $\quad Q = CV,\ Q_1 = C_1V_1,\ Q_2 = C_2V_2, = \cdots$

Then by substitution, $\quad CV = C_1V_1 + C_2V_2 + C_3V_3 + \cdots$

$$C = C_1 + C_2 + C_3 + \cdots \quad (187)$$

(*b*) *Series Arrangement.* — In this arrangement, the charge on all the plates but those connected to the battery is gotten by induction and since in all practical condensers the plates are exceedingly close together, we know that the induced charges are very nearly

equal to (though less than) the inducing charges on the plates attached to the battery. So we have:

$$Q = Q_1 = Q_2 = Q_3 = \cdots$$

and

$$V = V_1 + V_2 + V_3 + \cdots$$

By substitution,

$$\frac{Q}{C} = \frac{Q_1}{C_1} + \frac{Q_2}{C_2} + \frac{Q_3}{C_3} + \cdots$$

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots \tag{188}$$

125. The Energy of a Charged Condenser. — Consider a condenser to be charged by taking one electron at a time from one plate and putting it on the other. Each successive electron is taken from a plate of increased positive charge and taken to a plate of increased negative charge. So larger and larger amounts of energy have to be expended as the charge on the condenser increases. We may calculate the potential energy represented by the totally charged condenser as the sum of the work in taking each infinitesimal charge dQ across the plates. By definition of potential difference,

$$\text{Energy} = \text{Work done} = \int_0^Q V\,dQ = \int_0^Q \frac{Q}{C}\,dQ = \frac{1}{2}\frac{Q^2}{C},$$

or, using $Q = CV$, we obtain

$$W = \frac{1}{2}QV = \frac{1}{2}CV^2 = \frac{1}{2}\frac{Q^2}{C}. \tag{189}$$

From the definitions of the electrostatic units of Q and V, the work computed in the above equations will be in ergs. If Q is expressed in coulombs, V in volts, and C in farads, W will be in joules.

These units are discussed in § 175. It is sufficient to state at this time the following relations:

1 farad = 10^6 microfarads.
1 coulomb = 3×10^9 e.s.u. of quantity.
300 volts = 1 e.s.u. of potential.

126. The Loss of Energy upon Sharing of a Charge. — When two condensers, charged to different potential differences, have their like-charged plates connected together, the energy computed for the combined condenser is less than the sum of the separate energies before sharing. Experimentally, the sharing is always

accompanied by a spark through the air just before the plates come in contact. The loss in energy appears in the spark and the I^2Rt loss in the resistance of the circuit. The physical facts known are: (1) no electricity can disappear, so $Q_1 + Q_2$, before sharing, must equal $Q'_1 + Q'_2$ after sharing. (2) After sharing, the potential difference across each condenser must be the same; $V'_1 = V'_2$. From these facts the loss in energy may be computed. As a numerical example consider a condenser of 20 microfarads capacity charged to 1000 volts connected in parallel with one of 10 microfarad capacity charged to 100 volts. Originally

$$W_1 = \frac{1}{2}\frac{20}{10^6}1000^2 = 10 \text{ joules,}$$

$$W_2 = \frac{1}{2}\frac{10}{10^6}100^2 = 0.05 \text{ joules,}$$

and $$W_1 + W_2 = 10.05 \text{ joules.}$$

Also $$Q_1 = \frac{20}{10^6}1000 = 0.02 \text{ coulombs,}$$

and $$Q_2 = \frac{10}{10^6}100 = 0.001 \text{ coulombs.}$$

After sharing, the total charge is still 0.021 coulomb and the total capacity is 30 microfarads. The potential difference of the combination is

$$V'_1 = V'_2 = \frac{0.021}{\frac{30}{10^6}} = 700 \text{ volts.}$$

Then the total energy

$$W'_1 + W'_2 = \frac{1}{2}\frac{30}{10^6}700^2 = 7.35 \text{ joules}$$

Loss in energy = 2.70 joules.

The total energy may be obtained as follows without computing the final voltage:

$$W'_1 + W'_2 = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2}\frac{(0.021)^2}{\frac{30}{10^6}} = 7.35 \text{ joules.}$$

PROBLEMS

1. Two condenser plates of 100 cm². area are given opposite charges of 1000 e.s.u. The plates are 5 mm. apart. For a vacuum and for $k = 5$, find the values of Φ, E, and N. What potential difference exists between any point and one 2 mm. nearer one of the plates?

2. Prove that 1 farad = 9×10^{11} e.s.u. of capacity, 1 microfarad = 9×10^5 e.s.u. of capacity, 1 micro-microfarad = $\frac{9}{10}$ e.s.u. of capacity. If the plates of a radio air condenser are 2 mm. apart, how much area must the plates have to make the capacity 1 micro-microfarad? 1 e.s.u. of capacity?

3. In deriving Eq. 189, change the variable before integration so as to obtain $\frac{1}{2}\,CV^2$ directly.

4. Calculate the energy stored in a charged condenser which has a capacity of 140 mfd. and contains a charge of 0.077 coulombs.

5. Condenser A has a capacity of 20 mfd. and is charged to a difference of potential of 80 volts. Condenser B contains a charge of 0.00032 coulombs and capacity of 4 mfd. What is the combined capacity of A and B; (*a*) when they are connected in parallel; (*b*) when they are connected in series? What quantity of electricity might be called the total charge in the system in each case?

6. A given condenser has a capacity of 250 mfd. and contains a charge of 0.03 coulomb. The plates are connected to a second condenser and the potential of the first drops to 30 volts. How much energy did the first condenser lose?

7. Three condensers of capacities 10, 15, and 20 mfd. respectively are connected in parallel and a potential of 120 volts is applied across them. Compute (*a*) the charge on each condenser, (*b*) the charge in the system, (*c*) the energy stored in the system.

8. Solve Problem 7 when the condensers are connected in series.

9. A condenser of 5 mfd. capacity is charged from an 800 volt line and one of 10 mfd. from a 1200 volt line. The two charged condensers are then connected in parallel. What is the voltage across their terminals?

10. The plates of a 20 mfd. condenser are at a potential difference of 1000 volts. What is the energy stored? The charged condenser is then connected in parallel with an uncharged condenser whose capacity is 25 mfd. Calculate the energy stored in the second condenser.

11. A man has a supply of telephone condensers, the capacity of each being 2 mfd. He needs a capacity of about 0.08 mfd. How may he obtain it?

12. Show that if two condensers have capacities C_1 and C_2 and original potentials V_1 and V_2, the loss in energy on connecting them in parallel is

$$\frac{1}{2}\frac{C_1 C_2 (V_1 - V_2)^2}{C_1 + C_2}.$$

13. A condenser of capacity C is charged to a voltage V. It is then connected in parallel with an uncharged condenser of capacity C_1. The second condenser is removed and discharged and then connected to C again. If it is thus charged and discharged n times, show that the charge remaining on the first condenser is

$$CV\left(\frac{C}{C_1 + C}\right)^n.$$

Magnetism

127. Coulomb's Law and the Unit Pole. — The law of forces between magnetic poles was investigated by Coulomb. The law appears in the same algebraic form as does Coulomb's law for elec-

trostatic charges. The force varies directly as the strength of each pole and inversely as the square of the distance between them.

$$F = \frac{mm'}{\mu r^2}. \tag{190}$$

The proportionality constant μ is called the permeability of the medium.

A pole is said to have unit strength if when it is placed in a vacuum one centimeter from an equal and like pole there exists a mutual repelling force of one dyne. Thus, by definition, the permeability of a vacuum is unity.

Since Eqs. 168 and 190 are mathematically alike, we may expect to find and do find that the subjects of static electricity and magnetism have many theorems and methods of analysis in common. One important difference occurs in the magnetization of iron, nickel, cobalt, and a few alloys, namely, the value of μ is subject to large variations, its value depending upon the intensity of the magnetizing field.

128. Magnetic Field, Magnetic Induction, and Lines of Force. — A magnetic field is said to exist at a given point if a magnetic pole placed at that point is acted upon by a force. *The intensity of a magnetic field at a point is quantitatively defined as the force acting on a unit north pole placed at the point.* Its direction is the direction of that force. *If the force is one dyne, the field is said to be of unit intensity.* This unit of field strength is called the *gauss.** It is customary to picture a field by "lines of force." These lines indicate the intensity and direction of the field, the intensity being indicated by the number of such lines constructed through unit area normal to the field. Thus, a field intensity of four gauss is represented by four lines per sq. cm. (Fig. 99).

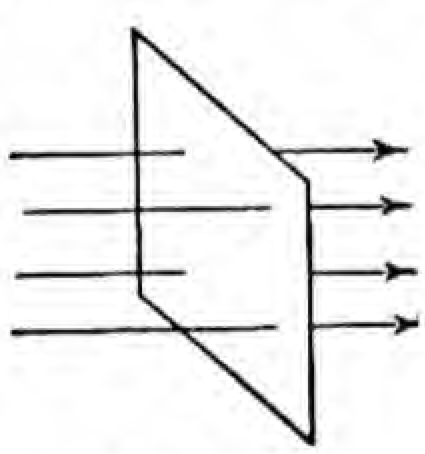


FIG. 99

* Very recently an international committee on electrical units has agreed to change the names of the units of magnetic induction, field strength, and reluctance (see § 184). Since all previous literature contains the old names, it seems advisable to use them throughout this text. The new usage is stated here for reference.

The unit of H, one dyne per unit pole, is to be called one *oersted*.
The unit of B, one line of induction per cm²., is to be called one *gauss*.
The unit of Φ, one line of induction, is to be called one *maxwell*.
The unit of reluctance, given no name, is simply one *gilbert per maxwell*.

It is thus seen that name of previous unit of reluctance has been shifted to the unit of field strength and likewise from the unit of field strength to that of induction, leaving the unit of reluctance (previously the *oersted*) as a derived unit with no name.

Just as in § 117, we see that if a pole of strength m is in a vacuum, the strength of the magnetic field at a distance r is m/r^2. The field is represented graphically by m/r^2 magnetic lines of force passing at right angles through each square centimeter of the sphere of radius r with the pole at its center. But when a medium of permeability μ is placed around the pole, then the intensity of the magnetizing field is changed to

$$H = \frac{m}{\mu r^2}.$$

So in order to associate a given number of lines with a pole of given strength we have to say that m/r^2 lines per square centimeter will still be drawn. These lines are called lines of induction. Their number per square centimeter is represented by B. So

$$B = \mu H. \qquad (191)$$

129. Torque on a Magnet Placed in a Field. Period of Oscillation. — In a similar logical procedure to that in § 115, using the facts of Eq. 190 and an agreement to measure magnetic field strength by the amount of force on a unit magnetic pole, and properly choosing units, we obtain

$$F = mH \text{ dynes}. \qquad (192)$$

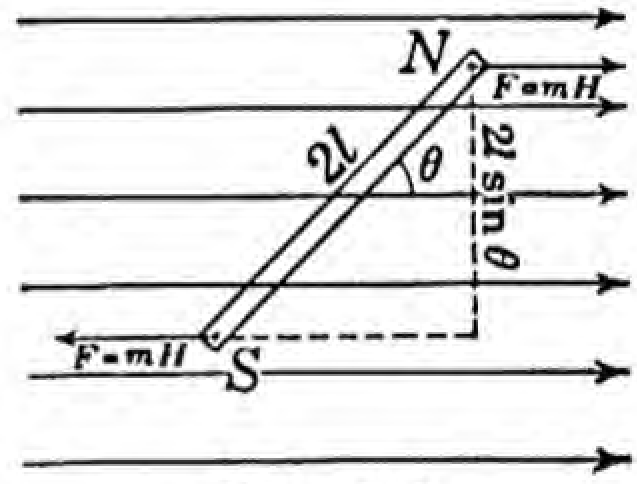


FIG. 100

If a magnet of length $2\,l$ is placed in a field as shown in Fig. 100, its axis making an angle θ with the field, there is a force of mH dynes on each pole. These forces, as shown, produce a couple whose moment is

$$2\,mHl \sin\theta = (2\,lm)\,H \sin\theta.$$

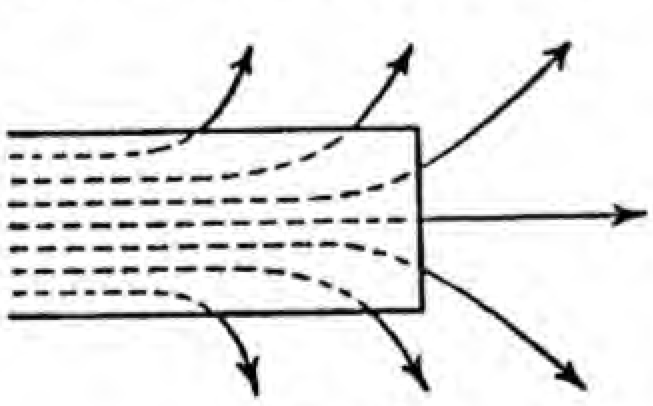


FIG. 101

Since the lines of force of a magnet when examined by a small compass needle or with iron filings appear somewhat as in Fig. 101, it is seen that the pole of a magnet is located not at a point but over a region. So the distance from the middle of the magnet to the pole is not determinable with accuracy. However, in the above expression, it is seen that H, θ, and the torque may all be measured accurately so an accurate value of the quantity $2\,lm$ may be calculated. This quantity $2\,lm$, the product of the pole strength and the distance between the

poles, is called *the magnetic moment of a magnet*, and is designated by the letter M. So

$$L = MH \sin \theta. \tag{193}$$

If $\sin \theta = 1$ and $H = 1$, then $L = M$, so the magnetic moment of a magnet is measured by the moment of force acting upon it when it is placed at right angles to a field of one gauss.

Now if a magnet is suspended freely from its center and then released, the torque acting upon it will cause it to turn into alignment with the field, but when it reaches the position where $\theta = 0$, it will have gained kinetic energy of amount $\int_\theta^0 L\, d\theta$ and so if no energy is lost in friction it will swing to the other side and oscillate between $+\theta$ and $-\theta$. The period may be determined by the laws of Simple Harmonic Motion if the angle of vibration is small because if θ is small then $\sin \theta$ is nearly equal to θ and $L = -HM\theta$. The negative sign is used because the return torque is always oppositely directed to the angular displacement. This is the condition for simple angular harmonic motion. The return torque per unit displacement is

$$L_0 = \frac{L}{\theta} = MH.$$

So from Eq. 92,

$$T = 2\pi \sqrt{\frac{I}{MH}}, \tag{194}$$

where I is the moment of inertia of the bar magnet. The value of I involves only mass and length measurements and T only seconds. Thus a purely magnetic quantity MH is determinable in absolute units of mass, length, and time.

130. The Intensity of the Magnetic Field about a Permanent Magnet. — In his early studies in magnetism Gauss computed the magnetic field strength at various points around a magnet. He used the letter A to represent a point along the axis of the magnet and B to represent a point on the perpendicular bisector of the length of the magnet. These letters are commonly retained.

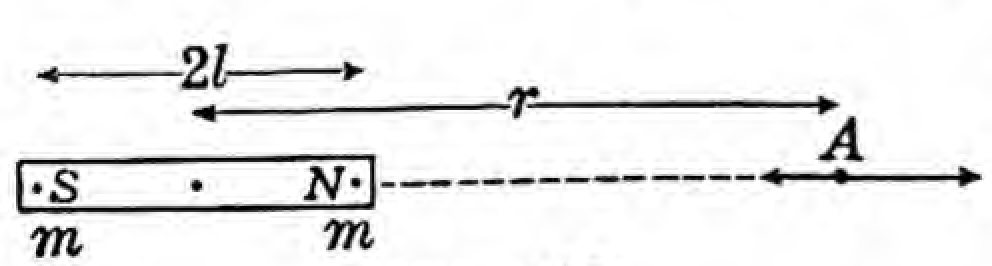


Fig. 102

(1) *The Gauss A Position.* — A magnet of pole strength m is considered (Fig. 102) and the field strength at the point A distant r from the center of the magnet is desired. For simplicity in

algebra, the length of the magnet is taken as $2l$. A unit north pole placed at A experiences two forces, one directed away from the magnet due to the repelling action of its north pole and a smaller force directed toward the magnet due to the attraction caused by its south pole. The resultant field at A is

$$H_A = \frac{m}{(r-l)^2} - \frac{m}{(r+l)^2} = \frac{4\,rlm}{(r^2-l^2)^2}. \tag{195}$$

If the point A is far enough from the magnet that l^2 is very small compared with r^2, then l^2 may be dropped and

$$H_A = \frac{4\,lm}{r^3} = \frac{2\,M}{r^3}. \tag{196}$$

It is to be noted that although l may be fairly large compared with r nevertheless l^2 may be small compared with r^2. Thus if l is $r/10$, l^2 is only one per cent of r^2. For such a case, if accuracy of greater than two per cent is desired, Eq. 195 must be used instead of Eq. 196.

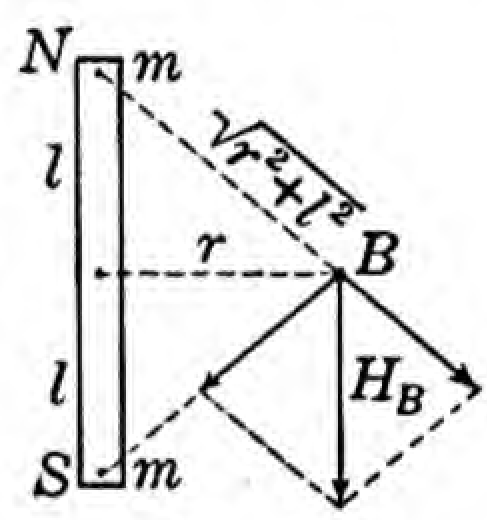


FIG. 103

(2) *The Gauss B Position.* — A unit north pole placed at B experiences two equal forces as shown in Fig. 103, each equal to

$$\frac{m}{r^2+l^2}.$$

The resultant is H_B. By similar triangles

$$\frac{H_B}{\dfrac{m}{r^2+l^2}} = \frac{2\,l}{\sqrt{r^2+l^2}}.$$

Then
$$H_B = \frac{2\,ml}{(r^2+l^2)^{\frac{3}{2}}}. \tag{197}$$

If, again, l^2 is small compared with r^2, then it may be dropped and

$$H_B = \frac{M}{r^3}. \tag{198}$$

It is noticed that $H_A/H_B = 2$. By carrying out these computations assuming an inverse nth power for the law of attraction, Gauss showed that $H_A/H_B = n$. By experimentally comparing the fields in the two positions, he proved $n = 2$ to considerable accuracy.

131. The Magnetometer. The Resultant of Two Fields at Right Angles. — An apparatus containing a magnet suspended

so as to rotate freely about an axis and arranged so that its deflections may be measured, is called a magnetometer.

Let a magnetometer be placed at P (Fig. 104) the axis of rotation being vertical (perpendicular to paper). In the figure, the magnetometer needle has been drawn large although in practice it must

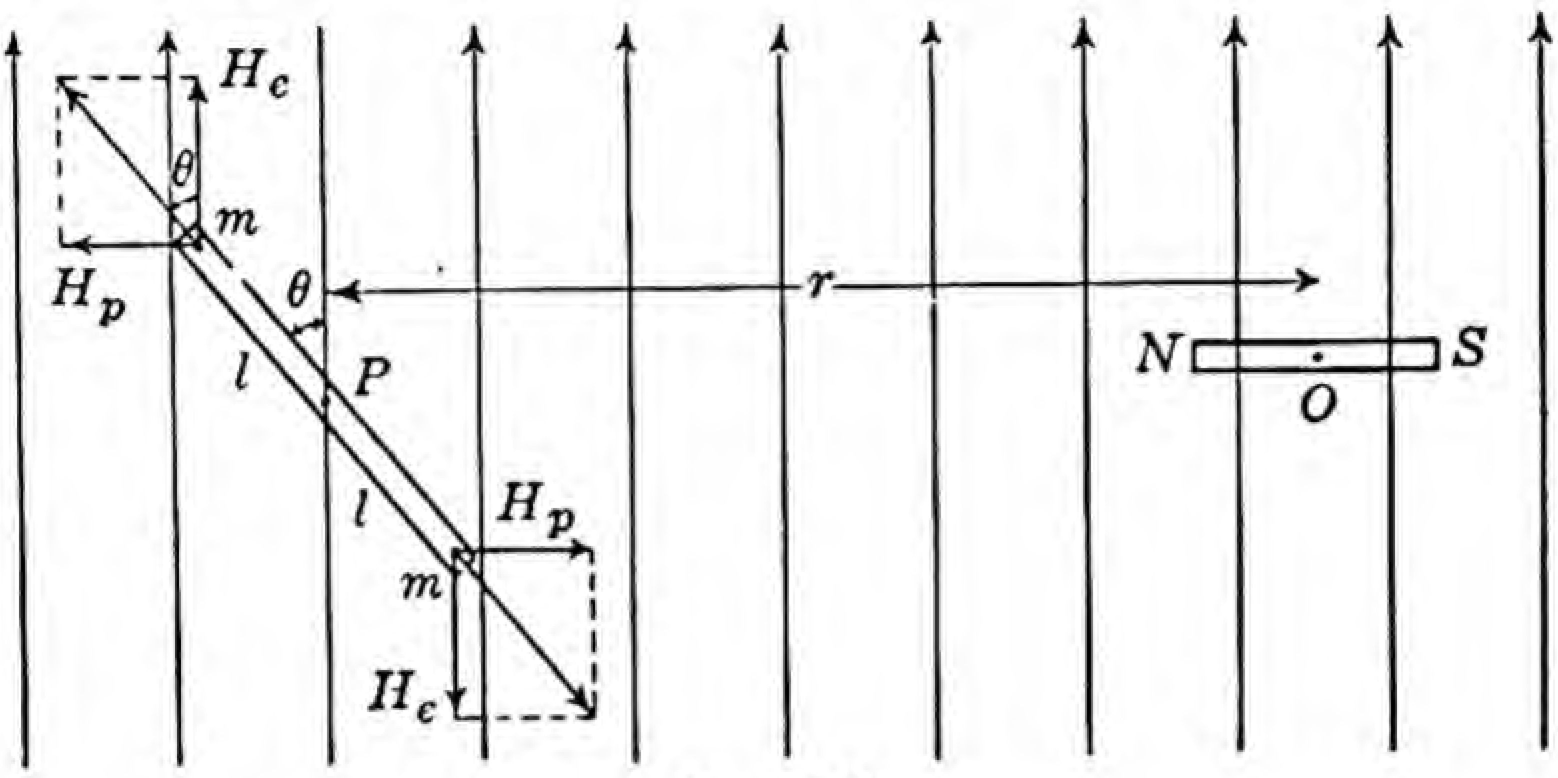


FIG. 104

be very small. The needle will take a position parallel to the earth's field whose strength is H_e. Directly east (magnetic east) of P at O, distant r, place a magnet of magnetic moment M or in any other manner create a field at P which will be directed east or west. Now the magnetometer needle will deflect and come to rest at some angle θ from the direction of the earth's field. The magnetometer is acted upon by two fields at right angles, one of intensity H_e directed north and the other of intensity H_p directed west due to the magnet placed as shown at O. The relations between θ and these two fields may be seen in two ways.

(1) The needle comes to rest at the angle θ where the torques due to the two fields are equal and opposite. Due to the earth's field there is a torque $H_e m\, 2\, l \sin\theta$. Due to the field at right angles to the earth's field there is a torque $H_p m\, 2\, l \cos\theta$. When equilibrium is reached,

$$H_e m\, 2\, l \sin\theta = H_p m\, 2\, l \cos\theta,$$

so

$$\frac{H_p}{H_e} = \tan\theta. \qquad (199)$$

(2) By definition, the needle at P takes the direction of the field at that point. The field at P is the resultant of the two fields. The intensities of these fields are vectors and their resultant must give the direction in which the needle comes to rest. By examination of the vector diagram, the relation shown in Eq. 199 is seen to hold.

132. The Absolute Measurement of M and H. — As was pointed out in § 129, it is possible to measure the product MH in terms of mass, length, and time. We shall now show how values of H and M may be obtained separately. Let the magnet be suspended at P (Fig. 104) the point where the existing magnetic field H_e is desired. (H_e may be the earth's field or any field existing at the point. No magnet is yet at O.) Its period when vibrating through a very small angle is measured. The same magnet is then placed at O and a magnetometer with a small magnetic needle is placed at P. The magnetometer needle deflects and Eq. 199 applies. Also the value of H_p is given by Eqs. 195 and 196. So

$$\tan \theta = \frac{2\,M}{r^3 H_e}$$

or more exactly

$$\tan \theta = \frac{2\,Mr}{(r^2 - l^2)^2 H_e}. \tag{200}$$

In this equation θ is the ratio of two lengths, and r and l are lengths, so the value of M/H_e is computed from measurements of length only. Now knowing the value of MH_e and M/H_e, both M and H_e are calculable.

Any such method by which the magnitude of an electrical or magnetic quantity may be determined solely by measurements of mass, length, and time, is called an absolute method.

133. Methods of Comparing Field Strengths. — After the above absolute method has been employed to obtain H at one station, there are several simple methods of determining the strengths of fields at other points by comparing them with the field already accurately measured.

First Method. — Studying Eq. 193, we see that if we use a permanent bar magnet of constant value of M, there are three variables, H, L, and θ. To compare two fields we must (*a*) keep θ constant and compare two values of L, or (*b*) vice versa.

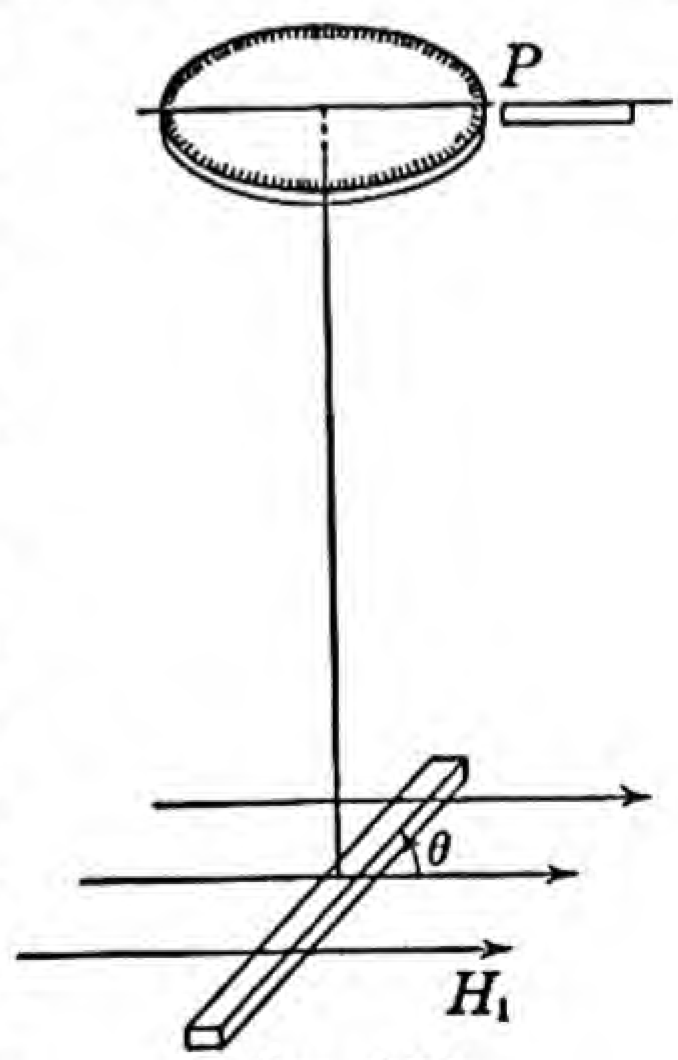


FIG. 105

(*a*) The magnet of moment M is fastened rigidly to an elastic suspension. The upper end of the suspension is fastened to a circular disk (Fig. 105). The disk is turned so that the suspension is under no torsion when the magnet is aligned with the field H_1.

The disk is then turned until the magnet has turned through an angle θ. The number of degrees, α_1, through which the disk is turned is noted by reference to a pointer P which is fastened to the main frame of the instrument. The number of degrees of twist in the suspension is then $(\alpha_1 - \theta)$. If k represents the torque necessary to twist the suspension 1°, then $k(\alpha_1 - \theta)$ is the torque required to turn the magnet through the angle θ. The instrument is then taken to the place where the field H_2 is desired. The disk is now turned through such an angle α_2 that the magnet is again turned through the angle θ. Applying Eq. 193 to each case, we have,

$$k(\alpha_1 - \theta) = L_1 = H_1 M \sin \theta,$$

and

$$k(\alpha_2 - \theta) = L_2 = H_2 M \sin \theta.$$

So

$$\frac{H_1}{H_2} = \frac{(\alpha_1 - \theta)}{(\alpha_2 - \theta)}.$$

(*b*) The student should describe the experiment and work out the equation for the case in which L is constant and θ the variable.

Second Method. — Let a given magnet be suspended and its period of oscillation be measured first at a place where the magnetic field has been accurately determined and then at a place where the value of field is desired. Applying Eq. 194, we have

$$\sqrt{\frac{H_1}{H_2}} = \frac{T_2}{T_1}, \quad \text{or} \quad \frac{H_1}{H_2} = \frac{T_2^2}{T_1^2}.$$

Third Method. — Again in Eq. 200 we have three variables H, r, and θ.

(*a*) The magnet of moment M is placed at O (Fig. 104) and the distance r is varied until the magnetometer needle has deflected through 10°, say. Then the whole apparatus is moved into the new field and the value of r again varied until the deflection is 10°. Then from Eq. 200 we have,

$$\tan 10° = \frac{2M}{r_1^3 H_1} = \frac{2M}{r_2^3 H_2},$$

or

$$\frac{H_1}{H_2} = \frac{r_2^3}{r_1^3}.$$

(*b*) The student should describe the experimental procedure in which r is constant and θ the variable and obtain the final formula.

PROBLEMS

1. A magnet placed in a field of 400 gauss intensity requires 1000 cm. dynes of torque to displace it 30° from the direction of the field. What torque will turn it at right angles to the field? What is the magnetic moment of the magnet?

2. A disk of radius 5 cm. and mass 10 gms. is fastened to a fine suspension (see Fig. 50). The period of this torsion pendulum is 10 seconds. The disk is removed and a magnet is fastened in its place. When placed in a field of 40 gauss, the top of the suspension has to be twisted through 170° in order to turn the magnet through 90°. What is the magnetic moment of the magnet?

3. A slender bar magnet is 8 cm. long. What per cent error is made in using Eq. 196 for the field at a point 50 cm. from the center of the magnet in the Gauss A position? How much error for the corresponding case for the Gauss B position? Explain the difference.

4. A very small magnet is suspended in the earth's field as shown in Fig. 104. The magnet at O produces a 10° deflection when $OP = 100$ cm. At what distance will the deflection be 20°? 30°? 60°?

5. The period of vibration of a horizontal magnet is 4.100 sec. at a place where the total intensity of the earth's field is 0.63 and the dip 73°. What is the total intensity of the field at a place where the period is 3.820 sec. and the dip is 60°?

6. A magnet is suspended by a fine suspension in a magnetic field of horizontal intensity of 0.140 gauss. Originally the suspension is unstrained. In order to turn the magnet 10° out of line with the field, the top of the suspension has to be twisted through 85°. What is the intensity of a field where a 100° twist of the top of the suspension produces a 15° turn of the magnet?

7. Make up a numerical problem to illustrate each case of the third method of comparing fields as described in § 133.

134. Gauss's Theorem. — In exactly the same proof as in § 118, we may replace q by m and k by μ and prove that *the total number of lines of induction across any closed surface is* 4π *times the number of unit poles enclosed within the surface.*

The student is expected to reproduce the modified proof.

135. Magnetic Potential. — In a manner similar to that in § 119, we define: *The magnetic potential at a point in a magnetic field is the work done in bringing a unit north pole from infinity to that point.*

Eq. 176 becomes, for the case of a pole of strength m at O (Fig. 94)

$$V = \frac{1}{\mu}\frac{m}{r}, \tag{201}$$

and for the potential difference between points distant a and b, respectively, from a pole of strength m we have

$$V_A - V_B = \frac{1}{\mu}\left(\frac{m}{a} - \frac{m}{b}\right). \tag{202}$$

Using the analogy with the electrostatic case, the student should give the definition of *unit magnetic potential difference* and also obtain the formula:

$$H = -\frac{dV}{dr}. \tag{203}$$

We shall consider the case analogous to that in Fig. 96. In the first figure we have the end faces of two permanent magnets, one a north pole and the other a south pole. There will be $4\pi m$ lines from one to the other. The number of lines of force per square centimeter is the same as the number of dynes of force on a unit pole placed between the poles. Let H_0 represent the intensity of this field. If the poles are moved closer so as to halve their separation, the field is still of strength H_0 although the magnetic potential difference is halved, since $\int H_0\, dr$ is halved. If, instead of moving the poles, we change the medium to one of permeability μ, the induction B (see § 128) remains the same, so therefore the field strength must decrease to a value $H = H_0/\mu$.

The case analogous to the second part of Fig. 96 is that of a long solenoid with an electric current flowing through it. The solenoid creates a field of constant strength,

$$H_0 = \frac{4\pi NI}{10\,L}$$

(see Eq. 217). If a very long magnetizable rod is placed inside of the solenoid, the induction increases to

$$B = \mu H_0.$$

The mechanism which accounts for these changes is given in the next section.

136. Molecular Theory and Demagnetizing Fields. — We may explain many of the phenomena of magnetism by assuming that there are many small magnets within the magnetizable medium, that they are ordinarily in haphazard arrangement exhibiting no magnetic effect, and that a magnetic field aligns some of these "molecular magnets" so that the material shows polarity. Fig. 106 represents a set of elementary magnets completely aligned. It is an experimental fact that when two magnets of equal pole strength have their opposite poles touching, the lines from the north pole of one enter the south pole of the other and produce no

external field. So we may consider in this figure that these small magnets when aligned have their lines as shown, the space between them being quite unaffected by their presence. Then the space in between the molecular magnets has passing through it only those lines of the original field of density H and therefore a unit magnetic pole placed in such a channel would experience H dynes of force.

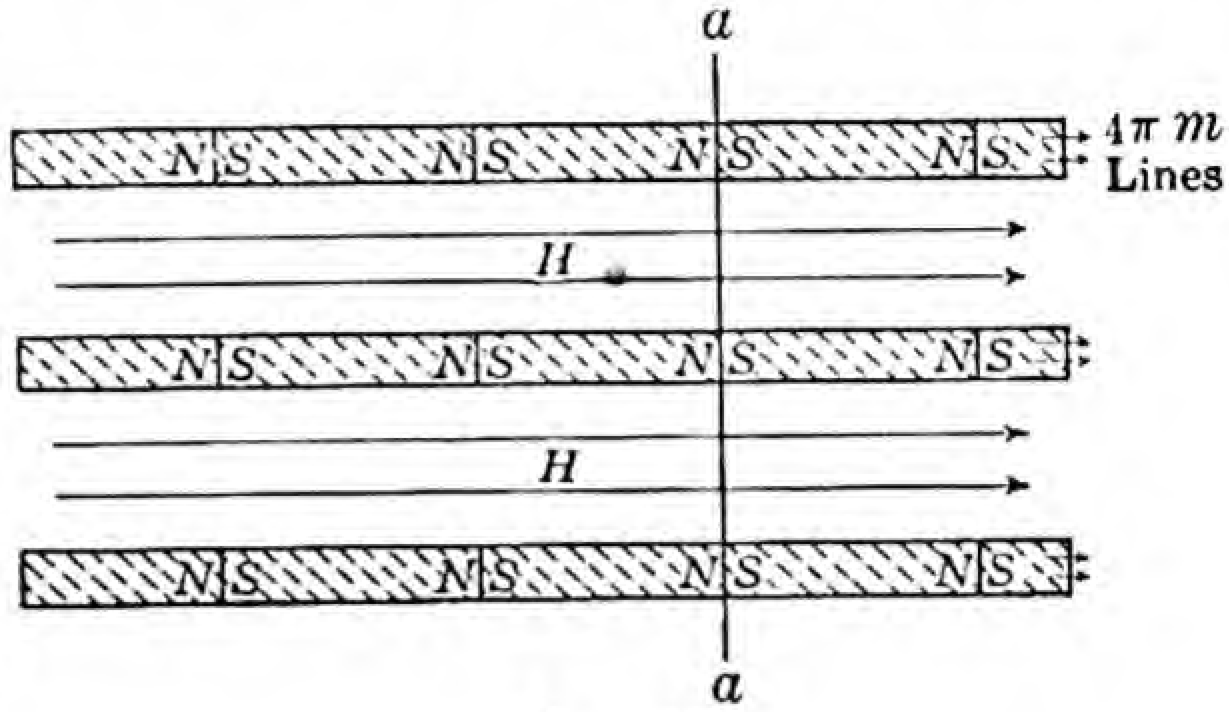


FIG. 106

Across a section aa there are lines of induction of number B per square centimeter, having the following relation,

$$B = H + 4\pi\mathcal{J} = \mu H, \tag{204}$$

where $\mathcal{J}$ represents the number of unit poles on each square centimeter of area of the plane through aa. $\mathcal{J}$ is called the intensity of magnetization.

Now if the magnetic material is cut through the plane aa and the molecular magnets separated so that their poles no longer touch (Fig. 107), then there appear the free poles at the surfaces and a unit test pole placed in this gap would experience the force of B dynes, the field strength in the gap being actually the value in air which is numerically equal to the induction in the iron. Practically, no such gap can be cut in a body without decreasing the original induction. It is to be carefully noted that the force on a unit pole *within* the body is at all times only H dynes.

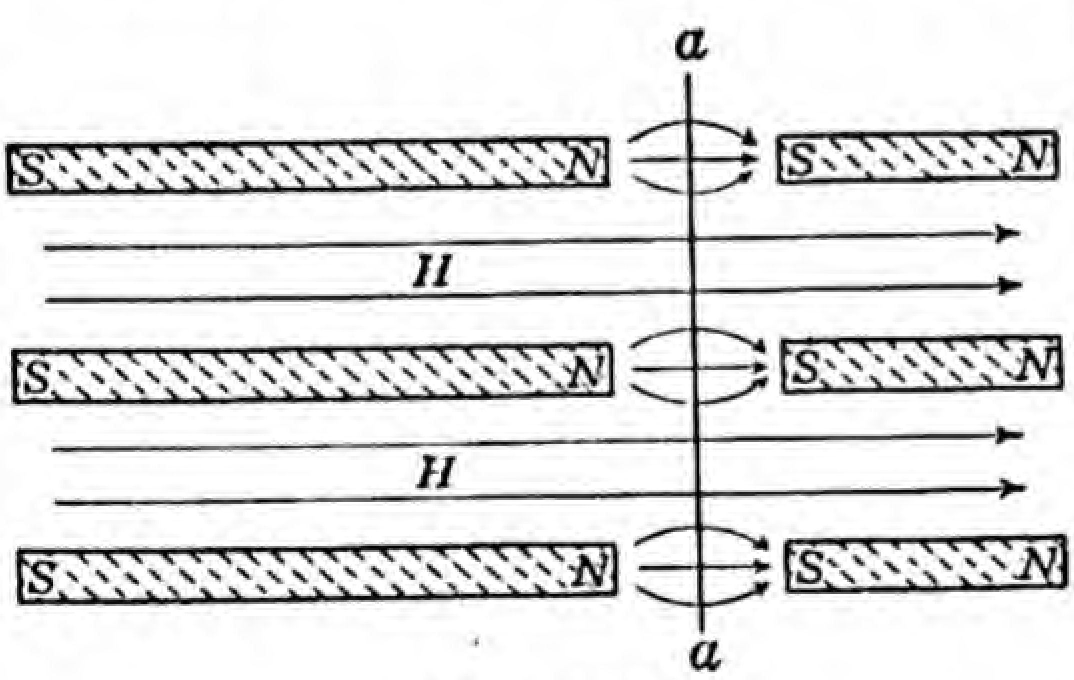


FIG. 107

The presence of a free pole (one without an equal opposite pole against it) always produces a demagnetizing effect. Consider any of the short magnets at the right side of Fig. 107. At any point inside of one of these magnets there is a field due to its poles, directed from the north pole to the south pole which is opposite to the applied field. Of course, this demagnetizing field must at all

times be less than the applied field or else the magnetization would never exist.

(1) *Case of a Ring Solenoid.* — If a magnetizing coil is wound uniformly on a ring of magnetic material, as shown in Fig. 108, the magnetizing field, H, is merely that produced by the current in the coil. There are no free poles in the magnet to produce any modification in this field. The induction is greater than H by the term $4\pi\mathcal{J}$. The slightest air gap in the ring has a remarkably large demagnetizing effect as will be computed in § 185.

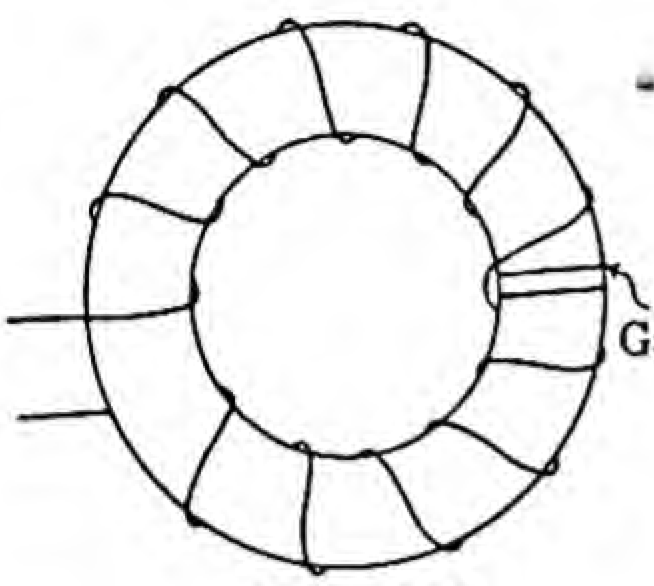


FIG. 108

(2) *Case of a Straight Iron Bar within a Straight Solenoid.* — Here the magnetizing field is modified by the free poles of the magnet and cannot be readily computed since it depends upon the dimensions of the iron bar as well as the state of magnetization of the iron. It should be carefully noted that H, the magnetizing field, is always the resultant of the field H', due to the coil, and the field P due to the free poles of the iron. Eq. 204 may be written for this case as

$$B = (H' - P) + 4\pi\mathcal{J}. \tag{205}$$

(3) *Case of Bunched Windings on an Iron Ring. Magnetic Leakage.* — The actual field due to the bunched windings is indicated by the dotted lines in Fig. 109. Only partial alignment of the molecules around the iron path is effected and free poles are produced on the surface of the ring. The resultant magnetizing field cannot be readily computed. It is somewhat less than in the case of uniform winding.

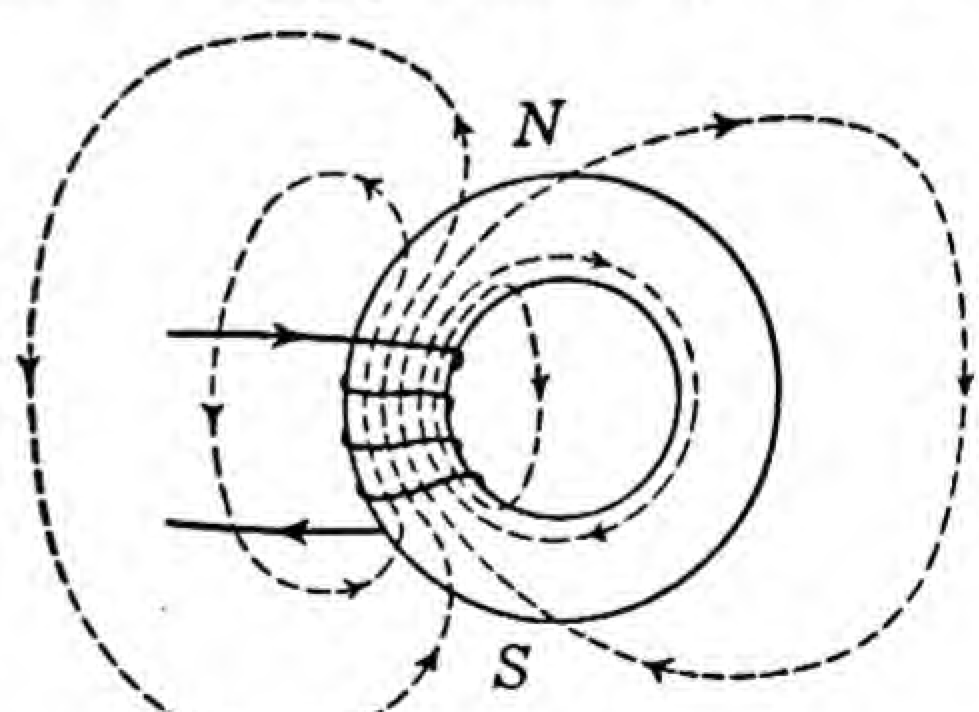


FIG. 109

The relation between B and H has been given two forms: $B = \mu H$ and $B = H + 4\pi\mathcal{J}$. The student should show that the following relation exists between the permeability and the intensity of magnetization,

$$\mu = 1 + 4\pi\frac{\mathcal{J}}{H} = 1 + 4\pi k, \tag{206}$$

where $k = \mathcal{J}/H$ is called the magnetic susceptibility.

137. Paramagnetism and Diamagnetism. — Most substances have a value of μ greater than unity. When placed near a strong magnetic pole, the end nearest the magnet becomes by induction a pole opposite to that of the magnet. They are thus attracted to a magnet of either polarity. Long rods of such substances when suspended freely in a magnetic field will align themselves with the field. These substances are, therefore, called paramagnetic. (The Greek prefix *para* means *beside, along with,* — as in *parallel.*)

A few substances behave in the opposite manner. The ends nearest a north pole become north poles by induction and vice versa, so they are repelled by either pole of a magnet. Rods of such substances will align themselves perpendicular to a non-uniform magnetic field. μ is thus less than unity. These substances are said to be diamagnetic. (The Greek prefix *dia* means *across* — as in *diameter.*)

The student should see from the equations in § 136 that k is positive for paramagnetic substances and negative for diamagnetic substances.

From the following tables, the values of μ for the diamagnetic substances are seen to be only very slightly less than unity (see Eq. 206). With the exception of iron, nickel, and cobalt, the values of the permeability of the paramagnetic substances are only slightly greater than unity. Although oxygen makes up only one fifth of the earth's atmosphere its relative larger paramagnetic property outbalances the weakly diamagnetic properties of hydrogen and nitrogen and causes air to be weakly paramagnetic. Liquefied oxygen is only weakly magnetic (although much more so than gaseous oxygen) but it will hang suspended between the horizontal pole pieces of a strong electromagnet. Substances are

MAGNETIC SUSCEPTIBILITIES

Special silicon-iron	4000.	Air	0.03×10^{-6}
Ordinary iron	200.	Hydrogen	-0.00017×10^{-6}
Nickel	23.	Nitrogen	-0.00040×10^{-6}
Cobalt	14.	Copper	-0.82×10^{-6}
Liquid Oxygen	324×10^{-6}	Zinc	-0.96×10^{-6}
Platinum	22×10^{-6}	Quartz	-1.07×10^{-6}
Magnesium	4×10^{-6}	Lead	-1.2×10^{-6}
Aluminum	1.8×10^{-6}	Silver	-1.5×10^{-6}
Tin	0.4×10^{-6}	Antimony	-5.61×10^{-6}
Oxygen	0.14×10^{-6}	Bismuth	-14.0×10^{-6}

called ferromagnetic if, like iron, their magnetic susceptibilities vary greatly with respect to the applied field strength. In the table are listed the maximum values for iron, nickel, and cobalt.

If we assume that in every atom there are electrons rotating in closed orbits, we may explain the magnetic behavior of matter. Each electron revolving in an orbit is equivalent to a circular current and thus each atom possesses a certain magnetic field about it. If these atoms are arranged at random, the fields all neutralize each other. A study of the forces on the rotating electron shows that when the material is placed in a magnetic field, the sizes of the orbits of the electrons are changed in such a way that a field is produced which opposes the applied field. Thus all substances will possess diamagnetic properties. But if in some materials the planes of the rotating electron are not rigidly fixed, they will tend to turn around in the applied field so that the field is increased. If this latter effect predominates over the small diamagnetic effect, then the substances are paramagnetic. In iron, nickel, and cobalt this effect of change in the plane of rotation is enormously larger than the effect of the change in size of the electron's orbit.

PROBLEMS

1. A piece of iron rod of 4 cm^2. cross-section is placed in a field of 200 gauss and its magnetization is such that each end has a pole strength of 5000 units. What is the permeability and induction?

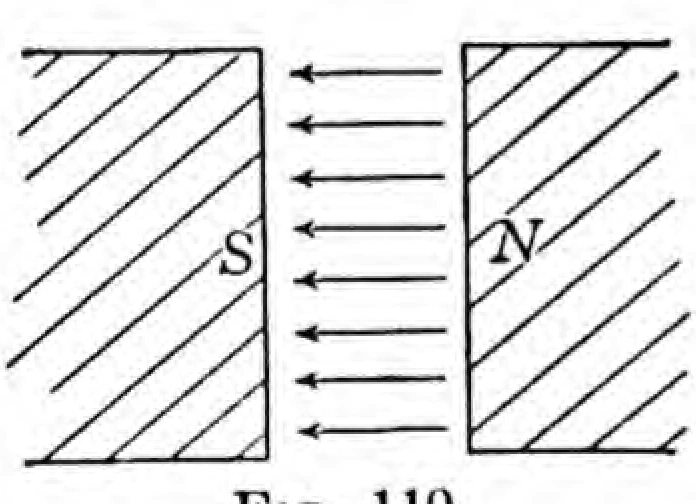


FIG. 110

2. The two pole pieces shown in Fig. 110 are 10 cm. apart. There are 5000 lines per cm^2. in the air gap. How much work is required to carry 10 unit north poles from the left-hand to the right-hand surface? What is the magnetic potential difference between the two surfaces?

3. Three north poles of 200 units strength are placed at the corners of an equilateral triangle with sides 10 cm. long. What is the value of the field strength and the potential at the center of the triangle? Solve the problem when one of the poles is changed for a south pole of the same strength.

THE MAGNETIC EFFECT OF AN ELECTRIC CURRENT

138. Side Push on a Current in a Magnetic Field. — When a current is placed in a magnetic field, there is an interaction between the magnetic field due to the current and the field in which the current is placed. The forces reach a maximum when the current is at right angles to the field and become zero when it is parallel to

the field. It will be remembered that the lines of force around a long straight wire carrying current are circles. Fig. 111 shows, superposed, the lines of force due to a current flowing perpendicularly into the plane of the paper and a uniform magnetic field directed at right angles to the current. Throughout this portion of the text, we shall use H_0 to represent the intensity of the field into which the current is placed and H to represent the field due to the current itself. At all points directly above the current the two fields have the same direction, their resultant being stronger than H_0. At all points directly beneath the current the two fields oppose each other. At A, near the current, $H > H_0$ and the field is directed toward the left. At a point B sufficiently far from the current, the value of H will be less than H_0 so that the resultant field will be directed toward the right. At some intermediate point N the two fields will be equal and opposite and produce a zero field. In Fig. 112 is shown the resultant of these two fields. The region where the lines are crowded together corresponds to the stronger portions of the field and conversely. Under these circumstances it is found that the current is acted upon by a force directed downward.

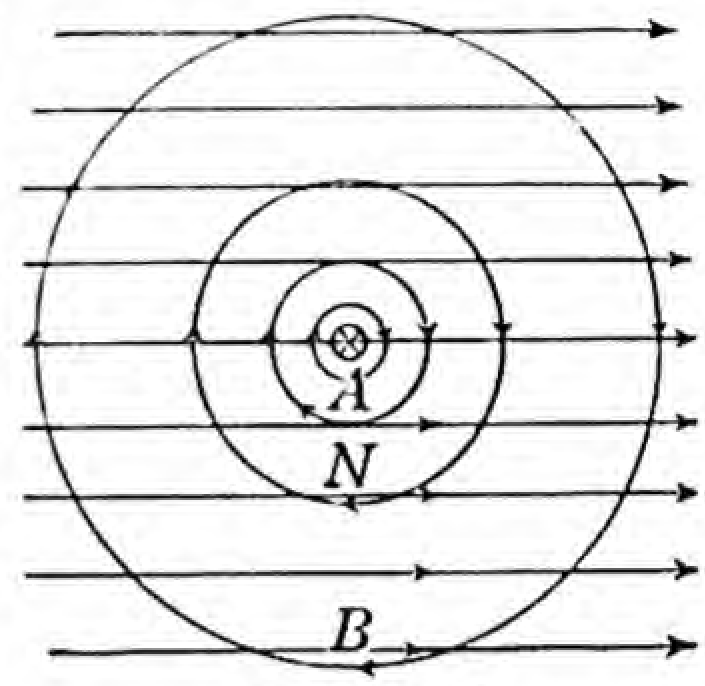


FIG. 111

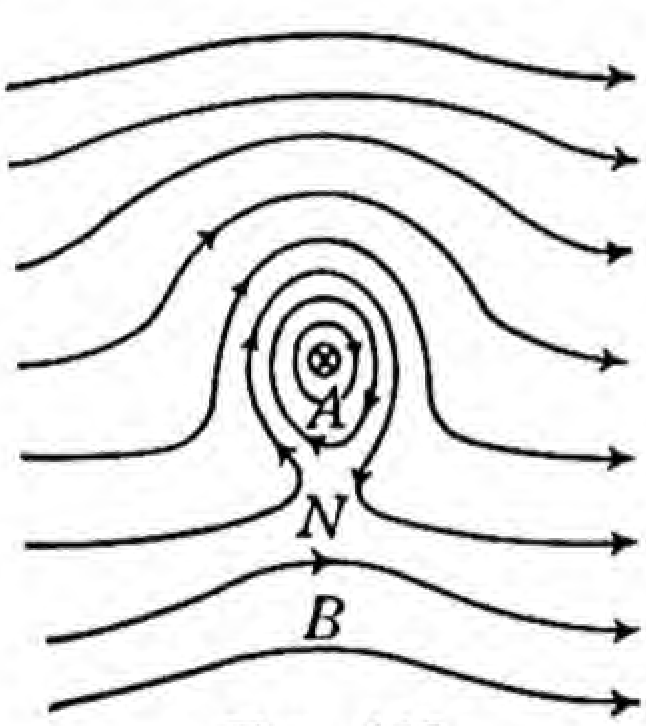


FIG. 112

In general, it is found experimentally that *when a current is placed in a magnetic field, it is acted upon by a force which is always perpendicular to the plane determined by the current and the lines of force which cross the path of the current.* Along this perpendicular *the force is directed from the stronger portion of the resultant magnetic field to the weaker portion.*

Experiment shows that the force acting on a straight wire (Fig. 113) making an angle θ with a magnetic field depends upon the current, the field strength, the permeability of the medium, the length of wire, and the angle, in a manner given by the expression

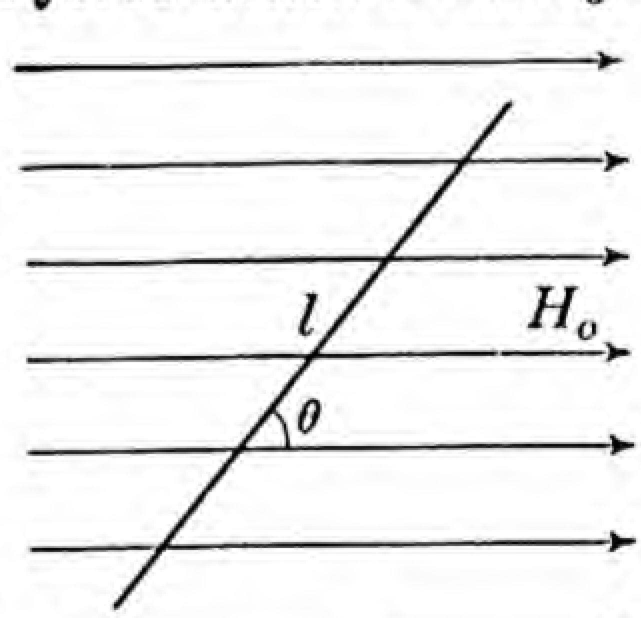


FIG. 113

$$F = K\mu H_0 il \sin \theta, \tag{207}$$

where K is a proportionality constant dependent upon the units only. This equation may also be written as

$$F = KBil \sin \theta, \tag{208}$$

where B is the magnetic induction.

139. The Magnetic Field Due to an Electric Current. — The magnetic effect at O (Fig. 114) due to the section dl of a current i is desired. Let a north pole of strength m be placed at O. This pole has a radial field about it the value of which in the region of dl is

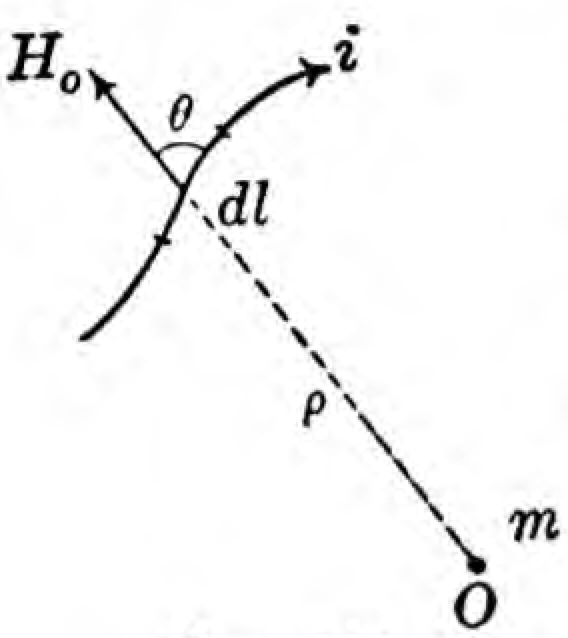


FIG. 114

$$H_0 = \frac{m}{\mu \rho^2}.$$

So the current shown in Fig. 114 will be acted upon by a force, directed perpendicularly out from the plane of the paper, whose magnitude, given by Eq. 207, is

$$dF = K\mu \frac{m}{\mu \rho^2} i\, dl \sin \theta = K \frac{im\, dl \sin \theta}{\rho^2}.$$

Now from Newton's Third Law of Motion, we know that if the pole m has a field which causes a certain force *outward* on the element of current, then the current must have a field at O which will cause the pole m to be pushed *inward* with an equal and opposite force.

Calling dH the field at O due to the current i in the element dl, the force on m is

$$m\, dH = dF = \frac{Kim\, dl \sin \theta}{\rho^2},$$

from which

$$dH = K \frac{i\, dl \sin \theta}{\rho^2}, \tag{209}$$

which is a result first obtained by Laplace and is called *Laplace's equation.*

To obtain the field due to any finite length of a wire, Eq. 209 must be integrated over the length of the wire considered.

Two very important facts are included in Eqs. 208 and 209. As an experimental fact it is stated in Eq. 208 that the force on a conductor in a magnetizable medium depends on the induction B. Derived from that fact, we see from Eq. 209 that the magnetic field surrounding a current is independent of the magnetic properties of the medium in which the current is placed.

140. Magnetic Field at the Center of a Circular Current. — Consider a current flowing through a circular coil of N turns, the radius of the circle being r. From a point at the center of the circle every element dl makes an angle of $\theta = 90°$ with its respective distance vector from that point. Integrating Eq. 209 so as to include the effect of all N turns,

$$H = \frac{Ki}{r^2}\int_0^{2\pi Nr} dl = K\frac{2\pi Ni}{r}. \tag{210}$$

141. The Electromagnetic Unit of Current and Potential Difference, the Ampere, the Coulomb, and the Volt. — The Electromagnetic Unit (e.m.u.) of current is defined in a manner so as to make $K = 1$ in the above equations. In Eq. 207 if we define i to be unity when $F = 1$ dyne, $\mu = 1$, $H_0 = 1$ gauss, $l = 1$ cm., and $\sin\theta = 1$, then $K = 1$. Thus *the e.m.u. of current is such a rate of flow of electricity through a wire in a vacuum at right angles to a field of one gauss that there will be a side thrust of one dyne on each centimeter of length of the wire.* Then K is unity in Eq. 210 and we may state a secondary definition: The e.m.u. of current is such a rate of flow of electricity in a circular coil of wire of one centimeter radius that there exists at the center of the circle a field of one gauss for every centimeter of length of the wire.

At the time of the definition of the e.m.u. of current, a smaller rate of flow was thought to be more practical for a unit, so the practical unit of current, *the ampere, is defined as one tenth of the electromagnetic unit of current.*

The e.m.u. of quantity of electricity is the quantity which passes in one second through a conductor in which the current is one e.m.u.

The coulomb is one tenth of the e.m.u. of quantity or is the quantity of electricity transferred each second by a constant current of one ampere.

If we use i for the current in e.m.u.'s and I for the current in amperes we now have the equations,

$$F = Bil\sin\theta = \frac{BIl\sin\theta}{10}\text{ dynes.} \tag{211}$$

$$dH = \frac{i\,dl\sin\theta}{\rho^2} = \frac{I\,dl\sin\theta}{10\,\rho^2}\text{ gauss.} \tag{212}$$

$$H = \frac{2\pi Ni}{r} = \frac{2\pi NI}{10\,r}\text{ gauss.} \tag{213}$$

Since the difference of potential between two points has been defined as the amount of work necessary to transfer a unit quantity

of electricity from one point to the other, then as soon as we have defined the unit quantity of electricity in any system of units we may then define the unit of potential difference. The unit potential difference in the electrostatic system has been defined in § 119.

The potential difference between two points is said to be one electromagnetic unit (1 e.m.u.) when one erg of work is required (or is done) in transferring one e.m.u. of quantity of electricity from one point to the other.

The potential difference between two points is said to be one volt when one joule (10^7 ergs) of work is required (or is done) in transferring one coulomb of electricity from one point to the other.

The student should prove that it follows from these definitions that one volt is equivalent to 10^8 e.m.u. of potential difference.

142. Magnetic Field at Any Point on the Axis of a Circular Current. — In Fig. 115 is shown the cross-section of the circular wire with its center at A. It is desired to find the magnetic field strength at a point O on the axis of the circle and distant d from it.

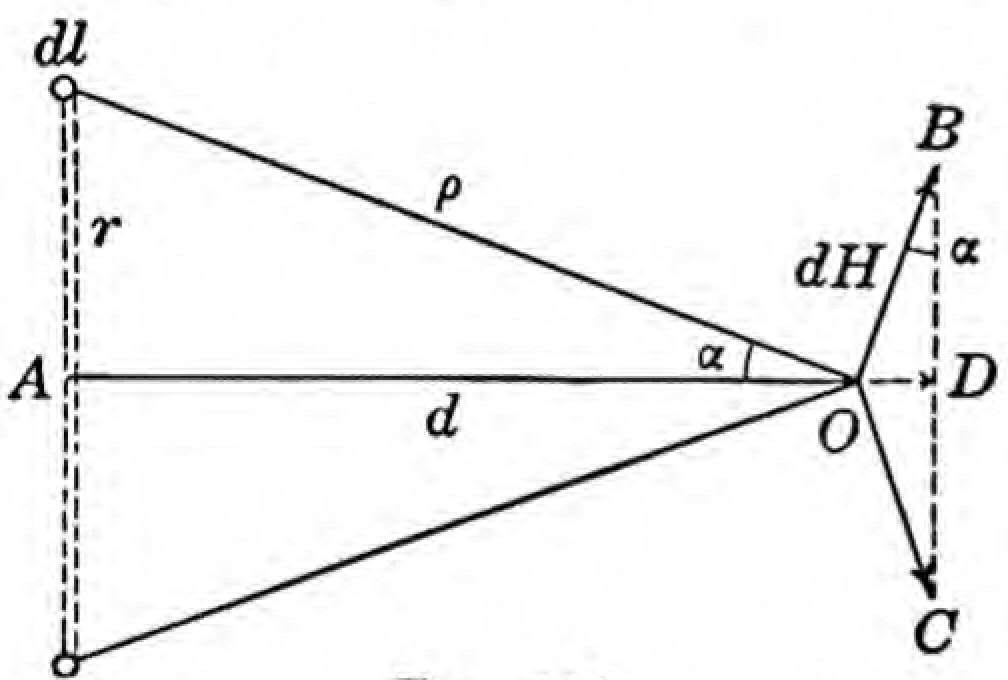


FIG. 115

Each current element dl is perpendicular to the radius vector ρ drawn from the point O. So θ in Eq. 209 is 90° and $\sin\theta = 1$.

Then $dH = \dfrac{i\,dl}{\rho^2}\sin\theta = \dfrac{i\,dl}{\rho^2}.$

The direction of this field for the element at the top of the circle (the current flowing out of the plane of the paper) is shown by the vector OB. In the element diametrically opposite dl, the current flowing into the plane of the paper causes a field equal in intensity to OB but directed along OC. The components DB and DC cancel each other, leaving the components OD to add up. Each current element thus has the component of its field perpendicular to the axis canceled by an element on the opposite side of the circle. Hence the resultant field strength will be found by summing up the components OD for all elements, *i.e.* $\int dH \sin\alpha$. Therefore, for N turns

$$H = \int_0^{2\pi Nr} \frac{i\,dl}{\rho^2}\sin\alpha = \frac{ir}{\rho^3}\int_0^{2\pi Nr} dl = \frac{2\pi N r^2 i}{(r^2 + d^2)^{\frac{3}{2}}}. \qquad (214)$$

When $d = 0$, this reduces to Eq. 213.

143. The Magnetic Field Due to a Straight Current of Infinite Length. — Every element dl (Fig. 116) produces a field at O which is directed into the plane of the paper. Laplace's equation gives

$$H = \int_{-\infty}^{+\infty} \frac{i\,dl \sin\theta}{\rho^2}.$$

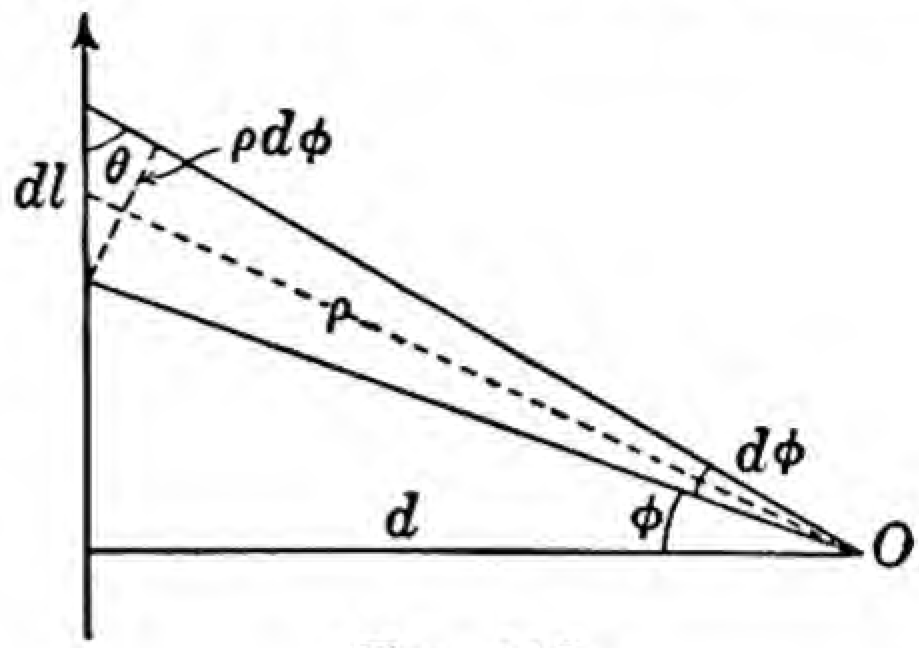


FIG. 116

The terms dl, ρ, and θ are all variables. Before the integration can be performed the expression must be reduced to one variable. The variable angle ϕ proves to be the most convenient. $dl \sin\theta = \rho\,d\phi$ and $\rho \cos\phi = d$, a constant. So

$$H = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{i \cos\phi\,d\phi}{d} = \frac{2\,i}{d} = \frac{2\,I}{10\,d}. \tag{215}$$

It is seen that although an infinite number of fields, dH, are added together, the sum of this infinite series is finite.

144. The Magnetic Field at the Center of a Long Solenoid. — The solenoid is simply a large number of circular currents so that the formula may be derived by properly summing up the expression in Eq. 214. Let L be the length of the solenoid (Fig. 117)

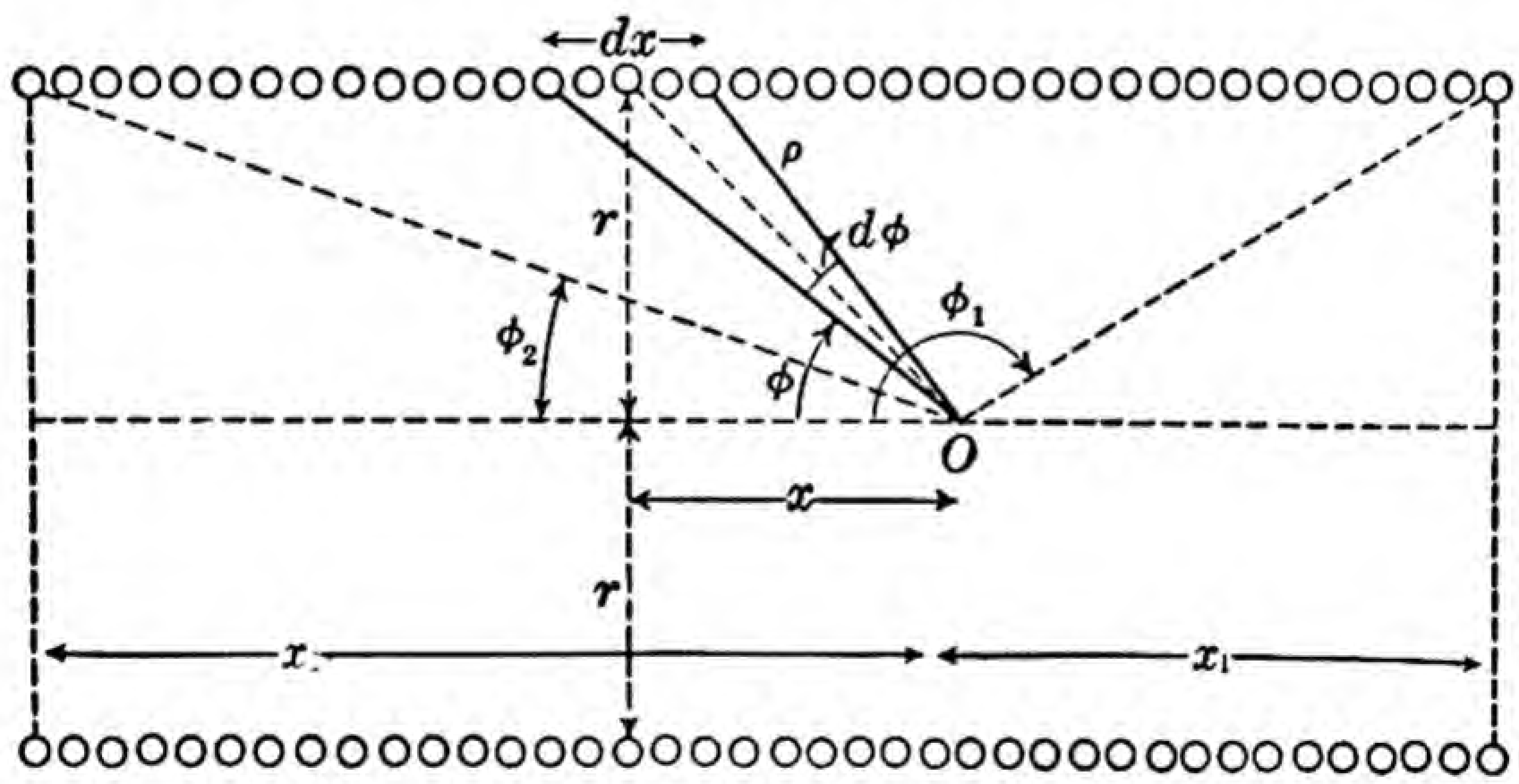


FIG. 117

and N be the total turns of wire. First we must find the field at a point O which is due to the current in a section of length dx which is distant x from the point O. The number of turns per unit length is N/L, and $N\,dx/L$ is the number of turns in the infinitesimal section of length dx. Substituting this latter expres-

sion for N in Eq. 214, we have the field at O due to the current in the section dx,

$$dH = \frac{2\,\pi r^2 N i\,dx}{L(r^2 + x^2)^{\frac{3}{2}}}.$$

Integrating this expression (see any table of integrals),*

$$H = \frac{2\,\pi N i x}{L(r^2 + x^2)^{\frac{1}{2}}}\Bigg]_{x_1}^{x_2} = \frac{2\,\pi N i}{L}\cos\phi\Bigg]_{\phi_1}^{\phi_2}$$

$$= \frac{2\,\pi N i}{L}(\cos\phi_2 - \cos\phi_1). \qquad (216)$$

For a very long solenoid $\phi_2 = 0$, $\phi_1 = 180°$, and

$$H = \frac{4\,\pi N i}{L}, \quad \text{or} \quad \frac{4\,\pi N I}{10\,L}. \qquad (217)$$

It may be shown by a more extensive proof that the field strength inside a very long solenoid is the same everywhere across its whole cross-section.

145. Work Necessary to Carry a Pole around a Current. — This work can be easily computed by taking the current in a long straight path as in § 143 because of the symmetry of the field. Let us carry a pole of strength m along the path of a circle of radius r whose center is on the line of the current. The magnetic field along the entire path is $H = 2\,i/r$. The force on a unit pole is therefore $2\,i/r$ dynes. The distance traveled in the direction of the force is $2\,\pi r$, so the work is

$$W = m\frac{2\,i}{r}2\,\pi r = 4\,\pi m i. \qquad (218)$$

The student should show that the amount of work done in moving a pole along the arc of a circle of any radius is a function of the angle subtended at the current and not of the radius. With this fact in mind we will show that the work in moving a pole once around a current is the same for any path whatsoever. Any irregular

* The expression may be integrated as follows: First note that $\rho d\phi = dx \sin\phi$, $\rho \sin\phi = r$, and $r^2 + x^2 = \rho^2$.
Substituting these expressions we find that

$$dH = \frac{2\,\pi N i}{L}\sin\phi\,d\phi.$$

Then

$$H = \int_{\phi_1}^{\phi_2} dH.$$

path AB can be broken up into small portions, as in Fig. 118. Over the portions that are directed along a radius from the current i, no work is required since there is no component of the force in that direction. Along the arcs described with the current as a center the work depends only on the angle subtended. By making the steps small enough the curve AB may be approached as a limit and so Eq. 218 holds for any closed path around the current.

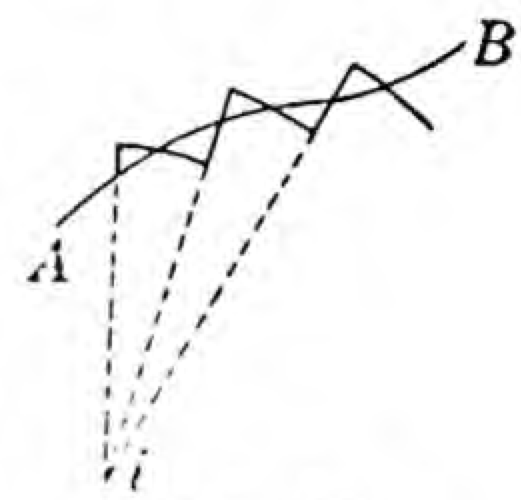


FIG. 118

Although Eq. 218 was derived for the case of an infinitely long straight current, it is possible to show that the formula holds for any closed path around any current.

146. Alternate Proof for the Magnetic Field in a Long Solenoid. — Consider the very long solenoid bent around so as to make a toroid (Fig. 119). Now carry a unit pole once around through the solenoid in a circular path of radius r. If the magnetic field inside the solenoid is H, then the force on the unit pole is H dynes. The total length of the solenoid is L. So $W = LH$. If there are N turns of wire, then the current enclosed in the path is Ni. Eq. 218 gives

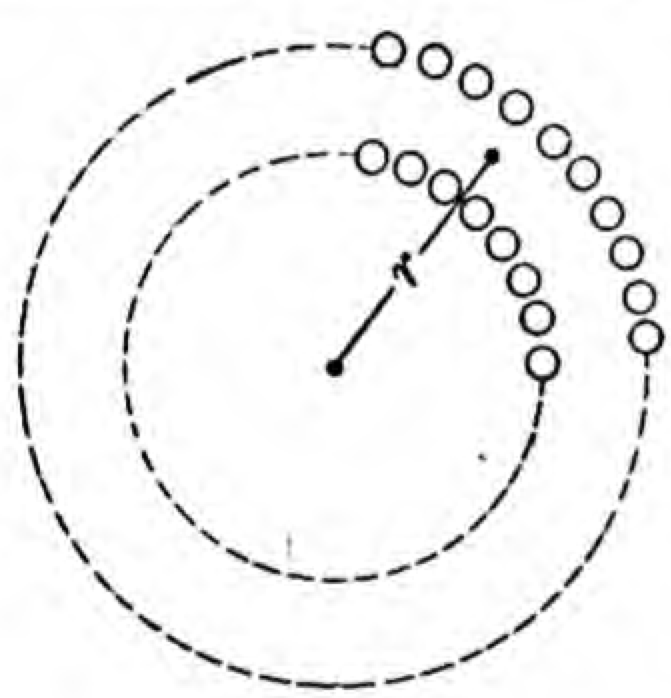


FIG. 119

$$W = LH = 4\,\pi Ni,$$

or

$$H = \frac{4\,\pi Ni}{L}.$$

This proof is not rigorous and furthermore does not offer any means of finding the field due to a short solenoid. The student should point out the assumptions made and the errors introduced at each step of the proof.

PROBLEMS

1. A circular coil of 50 turns has a diameter of 25 cm. Its magnetic field exerts a force of 2 grams weight on a magnetic pole having a strength of 300 e.m.u. placed at its center. What is the current in the coil?

2. A long straight horizontal wire running north and south carries a current of 30 amperes. A small compass placed below the wire points north-east. The horizontal component of the earth's magnetic field is 0.2 gauss. What is the direction of the current in the wire and how far is the compass from it?

3. A solenoid of 4 cm. radius and wound with 6 turns per cm. is 50 cm. long. Compute the exact field at its center when it carries a current of 5 amperes. Compute the field by the approximate formula.

4. A circular coil of 20 turns has a diameter of 30 cm. and carries a current of 3 amperes. It is placed vertically with its plane magnetically north and south at a place where the horizontal component of the earth's magnetic field is 0.2 gauss. In what direction will a compass point which is 20 cm. magnetically east of the coil in the line of its axis?

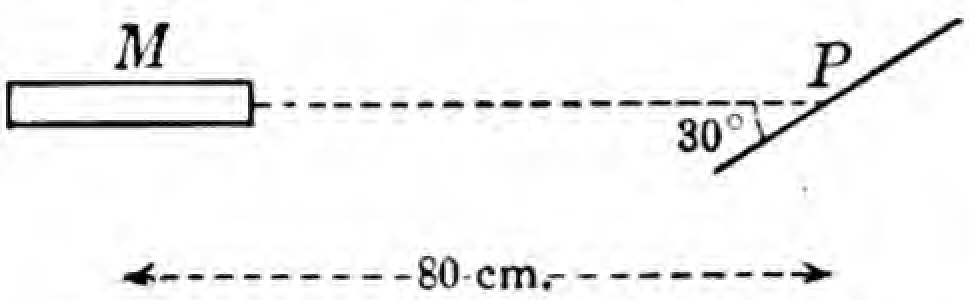


FIG. 120

5. The magnet M in Fig. 120 has a moment of 40,000 pole cm. A wire carrying 20 amperes crosses its axis at P, making an angle of 30° with the axis. What is the force per cm. exerted on the wire at P by the magnet?

6. At R (Fig. 121) a straight wire carrying 5 amperes intersects the plane of the diagram at right angles. Compute the work done in carrying a magnetic pole of 20 e.m.u. along (1) the arc AB; along (2) the radius BC; along (3) the straight line CD; along (4) the straight line AD.

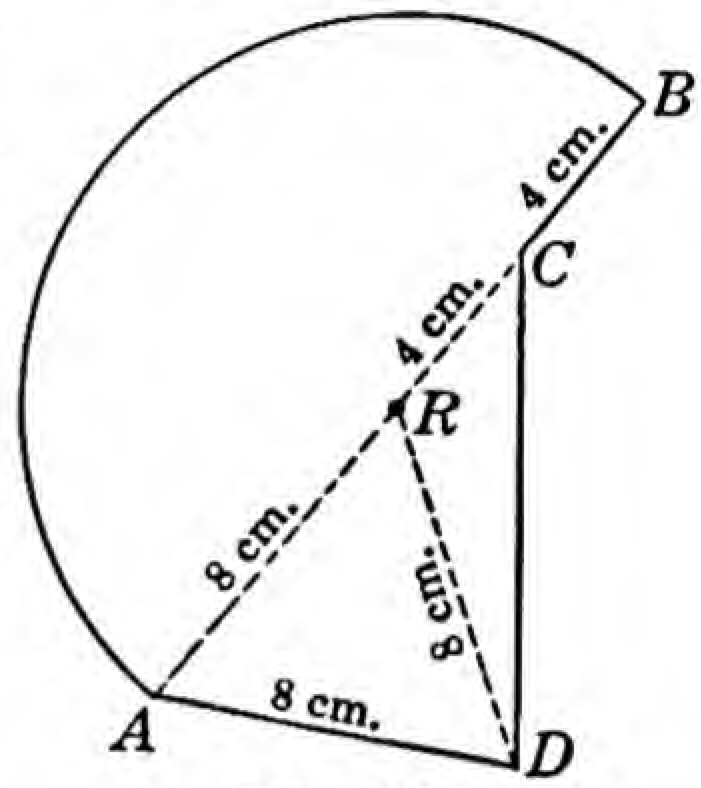


FIG. 121

7. A magnet is placed, as shown in Fig. 122, 20 cm. from a wire carrying current and at right angles to the wire. The current is 10 amperes and flows down (into the paper). Show by diagram the direction of the force on each pole and the resultant force on the magnet. Compute the magnitude of the resultant force.

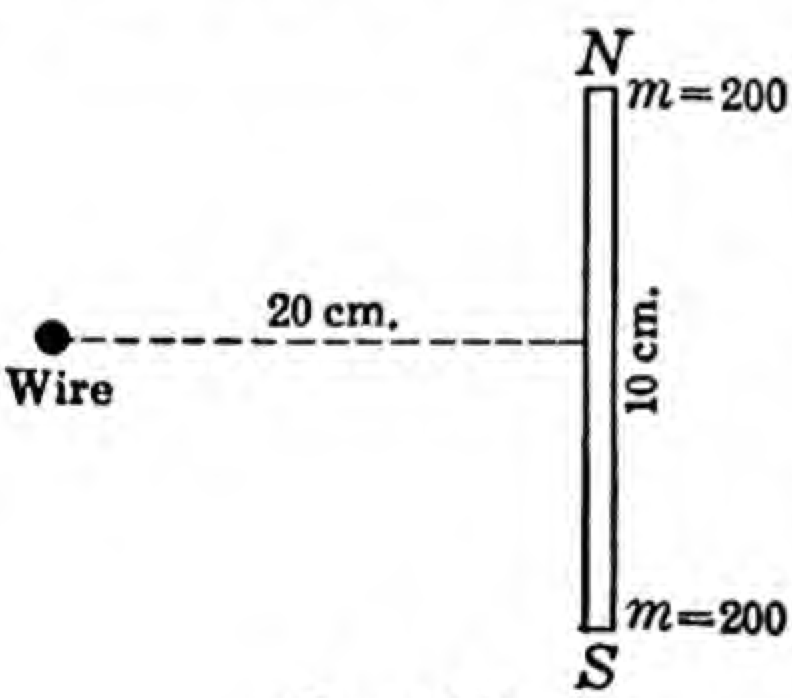


FIG. 122

8. A wrought iron rod 2 cm. in diameter and 100 cm. long is suspended by a string which passes over a frictionless pulley to a counter weight which balances it against gravity (Fig. 123). One end of the rod is flush with the end of a solenoid. The solenoid is 40 cm. long and has 400 turns. A current of 1.0 ampere flows in the coil and produces a field sufficient to cause magnetic saturation in the iron. The value of the permeability for this condition is 1000. Considering only the pole at the lower end, determine what additional weights must be placed on the pan to counterbalance the pull of the coil.

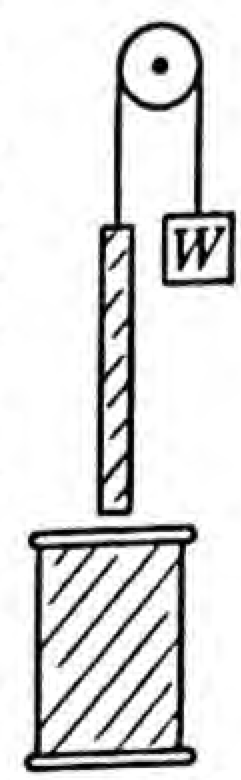


FIG. 123

9. A conductor 25 inches long lies in a slot between two pole pieces where there is a field of 10,000 gauss. What force acts on the conductor when 50 amperes flow through it?

147. Joule's Law; Resistance; Units. — Experiment shows that the heat developed in a conductor by an electric current is proportional to the time, to the square of the current, and to a quantity R, called the resistance, which depends upon the material,

the dimensions, and the temperature of the conductor. This is expressed by the following equation known as Joule's law, namely,

$$\text{Energy} = KI^2Rt, \tag{219}$$

where K is a constant dependent only upon the units used.

In order to make K equal to unity, *the resistance of a conductor is said to be one e.m.u. if one erg of energy appears as heat each second when the current is one e.m.u.* In the practical system, *the resistance is said to be one ohm, when one joule of heat appears each second when the current is one ampere.*

If a current flows steadily along a conductor of given resistance, it does so only because there is some driving force. This driving force is measured by the difference of potential and is the integral of the product of the electric field and the element of displacement, taken over the length of the conductor. We have seen from the definition of potential difference that a quantity of electricity, Q, passing from one point to another V units of potential lower, will do QV units of work. So Eq. 219 becomes

$$QV = I^2Rt. \tag{220}$$

Since $Q = It$, then

$$V = IR. \tag{221}$$

This is Ohm's law. Ohm's law and Joule's law are interconnected. From either one taken as the fundamental experimental fact, the other may be derived. From the definition of units of resistance as given above we may give the more common definitions in terms of Ohm's law:

If the ends of a conductor are maintained at a difference of potential of one e.m.u. and one e.m.u. of current flows through it, it is said to have one e.m.u. of resistance.

If the ends of a conductor are maintained at a difference of potential of one volt and one ampere of current flows through it, it is said to have one ohm of resistance.

The resistance of a piece of material varies directly as its length and inversely as its cross-sectional area, or

$$R = \rho \frac{l}{a}. \tag{222}$$

The value of the proportionality fact or ρ depends upon the chemical and physical properties of the material and is called the *resistivity* or *specific resistance* of the substance.

If R is expressed in ohms, l in cm., and a in cm²., ρ has the units of ohm cm. and is numerically equal to the number of ohms of resistance of a rod 1 cm². in cross-section and 1 cm. long.

In engineering work where, for all small wires of circular cross-section, the diameter is measured in thousandths of an inch, the computation of the areas is troublesome. So a new unit of area is used, called the circular mil, which is the area of a circle whose diameter is one mil (one thousandth of an inch). Since the area of a circle varies as the square of its diameter, then the area of a circle of diameter D mils is D^2 circular mils. The length l is expressed in feet and so the units of ρ are ohm circular mils/foot. In these units ρ is numerically equal to the resistance in ohms of a wire of 1 circular mil cross-section and 1 foot long.

The resistance of all materials is found to change more or less with temperature. Within rather wide temperature ranges, through which the materials do not change their molecular arrangement, their resistances may be expressed as a power series in the manner of Eq. 115.

148. Power. — From Eqs. 219 and 221 we get

$$\text{Power} = \frac{d}{dt}(\text{Energy}) = I^2R = \frac{V^2}{R} = VI, \tag{223}$$

provided I and R are constant. If I, V, and R are expressed in e.m.u., then the power is in ergs per second. If they are expressed in amperes, volts, and ohms respectively, then from previous definitions it is seen that the power is expressed in watts (joules per second).

In the reverse order we may express energy in terms of power. $\text{Energy} = \int \text{Power} \times \text{Time}$. So energy may be expressed as watt seconds (joules), watt hours, or kilowatt hours. In engineering work, 1 horsepower = 550 ft. lbs./sec. = 33,000 ft. lbs./min. and so energy might be expressed in horsepower seconds, horsepower minutes, or horsepower hours.

149. Ohm's Law for a Simple Series Circuit. — Let us now adapt Ohm's and Joule's laws to a simple series circuit.

Let the circuit possess a series wound generator and motor, and a resistance as in Fig. 124. The armatures and field windings of the generator and motor each have a definite, though small, resistance.

The potential difference across the terminals of a generator when it is supplying no current is called its Electromotive Force (e.m.f.), E_g. When it is supplying current, external mechanical energy of amount QE_g is required to be done on the generator for each quantity of electricity, Q, transferred around the circuit. When the motor is running, a back e.m.f., E_m, is produced, and for each quantity Q driven through it by the generator against its e.m.f. a quantity of external mechanical work QE_m will be done. Also during this process heat will be produced in the resistances R_g and R_m of the armature and field windings of the generator and motor respectively and in the external resistance R. We may apply the principle of conservation of energy to the case. So

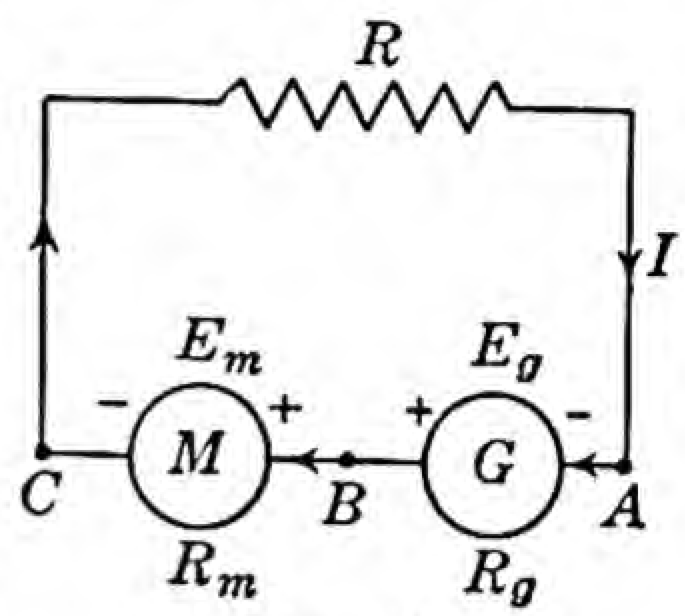


FIG. 124

$$QE_g = QE_m + I^2R_mt + I^2R_gt + I^2Rt. \tag{224}$$

Dividing by $Q = It$,

$$E_g = E_m + I(R_m + R_g + R),$$

or

$$\Sigma E = I\Sigma R, \tag{225}$$

where ΣE is an algebraic summation, an e.m.f. being considered positive when directed the same as the current and conversely.

The potential difference between different portions of the circuit can be obtained from Eq. 225. For the circuit from A to B, we select the only two terms represented in that portion, namely, the value of E_g and R_g and arbitrarily choose to let IR_g be a positive term. Thus.

$$V_{AB} = IR_g - E_g = - E_m - IR_m - IR. \tag{226}$$

With this convention about signs, in passing from any point A to a point B in the direction of the current, an IR drop is positive and an e.m.f. in the direction of travel is a rise in potential and therefore negative (because of its transfer to the opposite side of the equation). The right-hand member gives the potential difference as computed in going from A through the external circuit to B. The middle member is that same potential difference but it is computed for the path from A to B through the generator. This equation shows that, *when a current is caused to flow by a generator of any kind, the potential difference across the terminals of the generator is less than its e.m.f.* (measured on open circuit) *by the IR drop through its internal resistance.*

The potential drop across the terminals of a motor is obtained from Eq. 225 by collecting all terms involving E_m and R_m,

$$V_{BC} = IR_m + E_m = E_g - IR_g - IR. \qquad (227)$$

The right-hand member gives the value as computed from the external circuit, the middle member through the internal circuit. Thus *the potential difference across the terminals of a motor is equal to the sum of its back e.m.f. and the IR drop through its internal resistance.* Likewise,

$$V_{CA} = IR = E_g - E_m - IR_g - IR_m. \qquad (228)$$

The relations expressed in Eq. 225 are graphically represented in Fig. 125. The potential difference is indicated by vertical distances on the cylinder and the circumference represents the total resistance of the circuit. The potential difference from A to B of the circuit is indicated by the same letters on the cylinder. From A there is the rise in potential of amount of E_g volts and through the resistance of R_g there is a uniform decrease in potential per ohm of resistance. The difference in vertical height between points a and B represents IR_g volts. From the point B to C there is a drop in potential of E_m volts and IR_m volts. From C to A through the resistance R the potential drops uniformly by an amount IR. The diagram should be studied to see that Eqs. 225 to 228 are all represented by it.

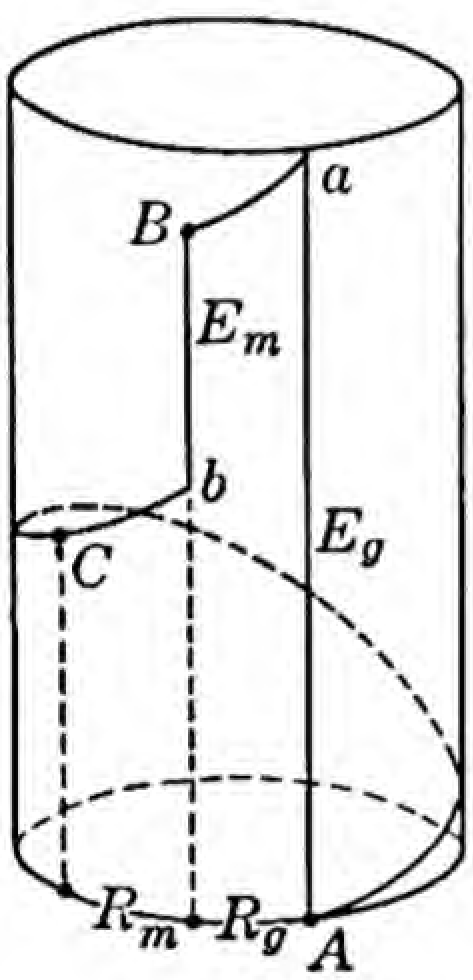


FIG. 125

The advantage of writing down the energy equation 224, is that no thought is necessary about the signs. Then algebraic transposition of terms gives the individual cases.

150. Sample Computations of Potential Differences in a Series Circuit. — To illustrate the method of applying Eqs. 225 to 228, let us consider the series circuit shown in Fig. 126. The e.m.f. and internal resistance of each cell is indicated as is also the value of each external resistance in the circuit. Throughout this book, the positive terminal of a battery is indicated by the longer of the two lines. The symbol ω is used to stand for "ohms." Applying Eq. 225,

$$8 + 1 - 4 = I(2 + 4 + 3 + 1 + 5)$$
$$I = \tfrac{1}{3} \text{ ampere.}$$

With a current of $\frac{1}{3}$ amp. flowing through the 4 ω resistance there is a potential difference of $\frac{4}{3}$ volts. The point B is $\frac{4}{3}$ volts higher than C. In traveling from the point B to point C, we say there is a drop in potential of $\frac{4}{3}$ volts but in traveling from C to B there is a rise of $\frac{4}{3}$ volts. We have chosen to call the drop in potential positive, hence IR is positive when traveling with the current and negative when against the current. In traveling from A to B, we have a drop of potential $\frac{1}{3} \times 2$ volts and a rise in potential of 8 volts. So

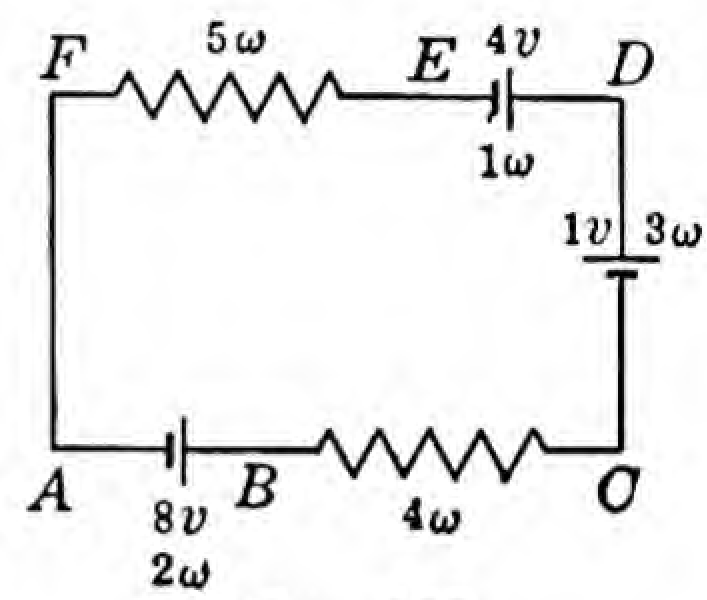


FIG. 126

$$V_{AB} = \tfrac{2}{3} - 8 = -\,7\tfrac{1}{3} \text{ volts.}$$

B is higher in potential than A by $7\frac{1}{3}$ volts. Likewise:

$$V_{CD} = -\,V_{DC} = \tfrac{1}{3} \times 3 - 1 = 0,$$

so C and D have the same potential,

$$V_{DE} = \tfrac{1}{3} \times 1 + 4 = +\,4\tfrac{1}{3} \text{ volts,}$$
$$V_{EF} = \tfrac{1}{3} \times 5 = \tfrac{5}{3} \text{ volt.}$$

Hence the sum of all the potential differences around the whole loop is

$$V = \tfrac{1}{3} \times 2 - 8 + \tfrac{1}{3} \times 4 + \tfrac{1}{3} \times 3 - 1 + \tfrac{1}{3} \times 1 + 4 + \tfrac{1}{3} \times 5 = 0.$$

This equation states that there is as much rise in potential as there is drop in potential so that the algebraic sum is zero.

151. Resistances in Series and Parallel. — The student is expected to draw diagrams for the following proofs and to explain the reason for each step.

For resistances in series the known physical facts may be stated in the following equations:

$$V = V_1 + V_2 + V_3 + \cdots$$

and
$$I = I_1 = I_2 = I_3 = \cdots$$

By the definition of resistance,

$$V = IR, \quad V_1 = I_1R_1, \quad V_2 = I_2R_2, \cdots$$

By substitution,

$$IR = I_1R_1 + I_2R_2 + I_3R_3 + \cdots$$

Hence
$$R = R_1 + R_2 + R_3 + \cdots \tag{229}$$

For resistances in parallel we know the following facts:

$$V = V_1 = V_2 = V_3 = \cdots$$

and

$$I = I_1 + I_2 + I_3 + \cdots$$

By substitution,

$$\frac{V}{R} = \frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3} + \cdots$$

Hence

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots \tag{230}$$

A very common case occurs when there are just two resistances in parallel. It should be proved and remembered that for this case

$$R = \frac{R_1R_2}{(R_1 + R_2)}.$$

Two resistances in parallel *always* have a resistance *less than the lesser* one of the two.

152. Divided Circuit. — When a current I divides and passes through two parallel branches, the fraction of that current which flows through each branch is easily computed. For this case, as shown in Fig. 127, we know that

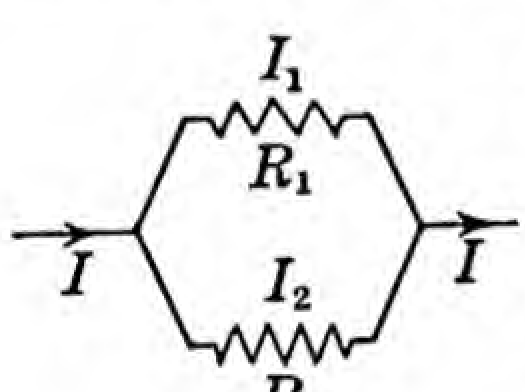


FIG. 127

$$I_1R_1 = I_2R_2,$$

and

$$I = I_1 + I_2. \tag{231}$$

These two equations contain three variables, I, I_1, and I_2. They may be combined into one equation containing any two of the variables. The equations we are interested in are, one containing I_1 and I, and one containing I_2 and I. The first of these is obtained by eliminating I_2 from Eqs. 231 and the second by eliminating I_1. We thus obtain:

$$I_1 = I\frac{R_2}{R_1 + R_2}, \quad \text{and} \quad I_2 = I\frac{R_1}{R_1 + R_2}. \tag{232}$$

In either case there is a definite fraction of the main current which passes through a given one of the resistances. For either one of the resistances, that fraction is the resistance of the opposite branch divided by the sum of the resistances.

153. Kirchhoff's Laws. — There are many electrical circuits that are not simple combinations of series or parallel branches. For these circuits, Kirchhoff states two laws

I. At every branch point,

$$\Sigma I = 0. \tag{233}$$

II. For each loop of the circuit,

$$\Sigma E = \Sigma IR. \tag{234}$$

(The algebraic summation is understood.)

The first law states that, when a steady state has been reached, the electricity cannot accumulate at a point; as much electricity must flow away from a branch point as flows into it.

The second law states that starting at any point of a circuit and traveling once around any loop, one finds that the algebraic sum of the e.m.f.'s of all cells, motors, and generators must equal the algebraic sum of the potential drops along the resistances.

The following procedure is usually followed in applying the laws (see Fig. 128):

1. Assume a different current in each branch of the circuit. Letter each current and indicate, with an arrow, an arbitrary direction for each.
2. Apply Law I to each branch point.
3. Apply Law II to each loop.

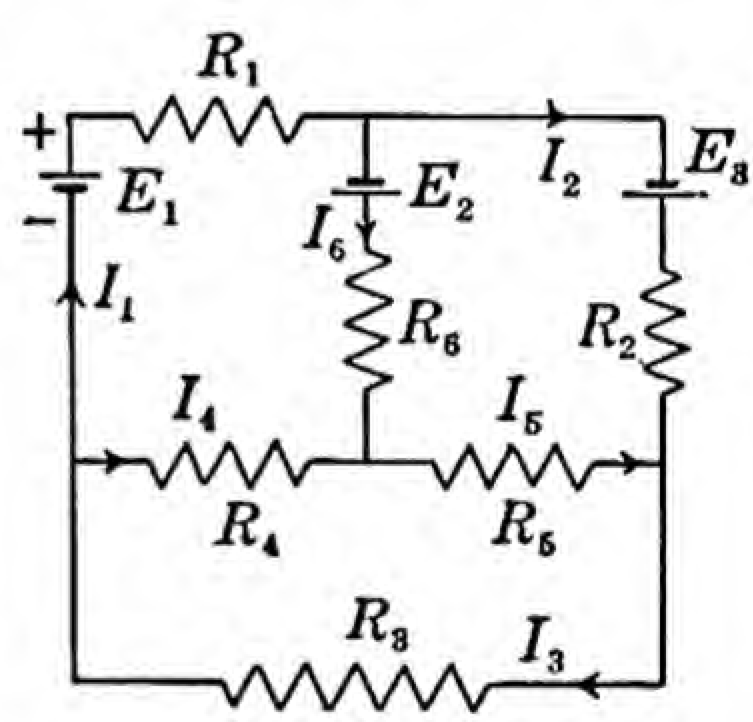


FIG. 128

Thus at the upper center point of Fig. 128,

$$I_1 - I_6 - I_2 = 0.$$

Hence $\qquad I_6 = I_1 - I_2$

and instead of using I_6, its equivalent $(I_1 - I_2)$ is used. Similarly

$$I_4 = I_3 - I_1 \quad \text{and} \quad I_5 = I_3 - I_2.$$

Replacing I_4, I_5, and I_6 by these values, it is seen that the currents in the branches can be expressed in terms of just I_1, I_2, and I_3. In general, the application of Law I always reduces the number of variable currents to the number of separate loops of the circuit.

The second law is applied separately to each loop of the circuit. One must remember, as explained for the use of Eq. 226, that E is positive when directed in the chosen direction of traversing the loop and that an IR drop is positive when the current is flowing in the direction of travel, and conversely for the negative signs. The clockwise direction is usually taken for traversing the loop. Using the results of the application of the first law, we thus obtain:

For the upper left loop,

$$E_1 + E_2 = I_1R_1 + (I_1 - I_2)R_6 - (I_3 - I_1)R_4.$$

For the upper right loop,

$$-E_2 + E_3 = I_2R_2 - (I_3 - I_2)R_5 - (I_1 - I_2)R_6. \tag{235}$$

For the lower loop,

$$O = (I_3 - I_1)R_4 + (I_3 - I_2)R_5 + I_3R_3.$$

Rearranging terms —

$$\begin{aligned} E_1+E_2 &= I_1(R_1+R_4+R_6) && -I_2R_6 && -I_3R_4, \\ -E_2+E_3 &= -I_1R_6 && +I_2(R_2+R_5+R_6) && -I_3R_5, \\ O &= -I_1R_4 && -I_2R_5 && +I_3(R_4+R_5+R_3). \end{aligned} \tag{236}$$

The value of I_2 is found by solving the three equations simultaneously, which is the same as evaluating the determinant:

$$I_2 = \frac{\begin{vmatrix} R_1 + R_4 + R_6 & E_1 + E_2 & -R_4 \\ -R_6 & -E_2 + E_3 & -R_5 \\ -R_4 & O & R_4 + R_5 + R_3 \end{vmatrix}}{\begin{vmatrix} R_1 + R_4 + R_6 & -R_6 & -R_4 \\ -R_6 & R_2 + R_5 + R_6 & -R_5 \\ -R_4 & -R_5 & R_4 + R_5 + R_3 \end{vmatrix}} \tag{237}$$

Likewise for I_1 and I_3. If any current comes out negative, then the current flows in the opposite direction to that assumed. The current I_6 is computed as $I_1 - I_2$ and its direction is that assumed if this difference is positive, and likewise for the current in the other two linking resistances.

154. Maxwell's Circuital Current Method for a Network. — In using the Kirchhoff method, Maxwell noticed that the current in the conductors that linked two circuits appeared as the difference of two currents (see the coefficients of R_4, R_5, and R_6 in Eqs. 235), which suggested that, mathematically, two currents might be considered to be flowing in opposite directions through the resistance at the same time. The rules for Maxwell's method are:

1. Consider that a separate current flows (preferably clockwise) through each loop of the network.
2. Consider that the difference between the currents of two neighboring circuits flows through the linking resistances and write down the equations for Kirchhoff's second law.

Thus for Fig. 129, which is the same as Fig. 128 except for the circuital currents, we have a set of equations which are identical with Eqs. 235. This method outwardly eliminates the use of Kirchhoff's first law. Purposely, the currents in the outer branches of Figs. 128 and 129 have been lettered the same and so the values of I_1, I_2, and I_3 must be the same by both methods.

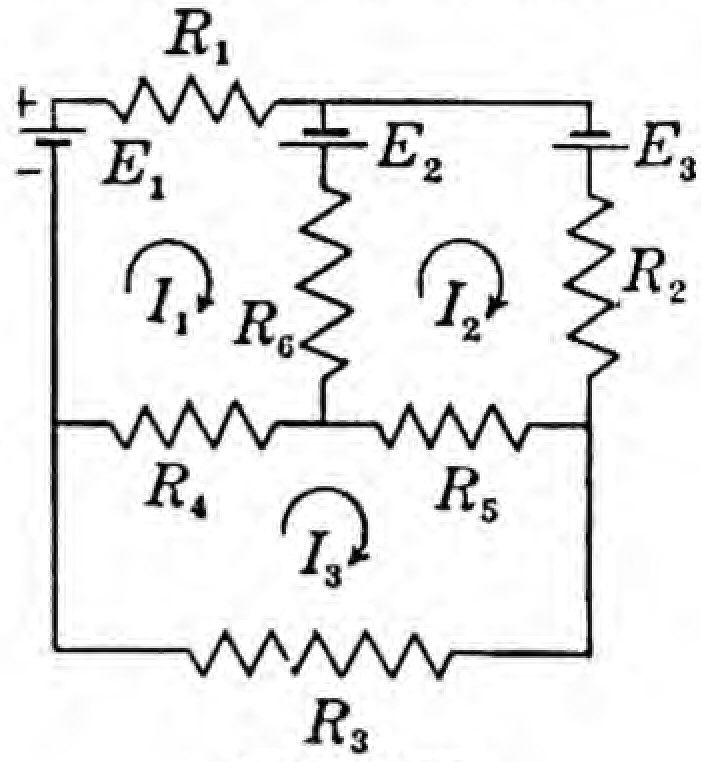


FIG. 129

If the Eqs. 236 are examined carefully, the following rules are seen to apply for the Maxwell method:

1. Indicate the circuital currents, one in each loop.
2. Write down one equation for each loop according to the following scheme:

Multiply the circuital current by the sum of all resistances in the loop and subtract the products of each linking current by the resistance through which it flows. Equate this to the algebraic sum of the e.m.f.'s (considering each positive when it tends to send current in the direction of the circuital current, in agreement with our former convention of signs).

PROBLEMS

1. With an external resistance of 9 ohms, a certain battery gives a current of 0.43 ampere. When the external resistance is increased to 32 ohms, the current decreases to 0.2 ampere. What is the resistance and the e.m.f. of the battery?

2. The resistivity of annealed copper is 1.724×10^{-6} ohm cm. What is the value of the resistivity in ohm circular mils/ft.?

3. Twenty-five feet of a piece of wire has a resistance of 10.7 ohms. If the diameter is 5 mils, what is its resistivity?

4. An electric resistance furnace wound from a single strip of wire is designed to operate on a 110 volt line. It is later desired to run the furnace at a higher temperature. Should the resistance wire be made shorter or have an extra section added to its length? What would be the effect on the total resistance and the temperature of the furnace of adding another similar resistance in parallel?

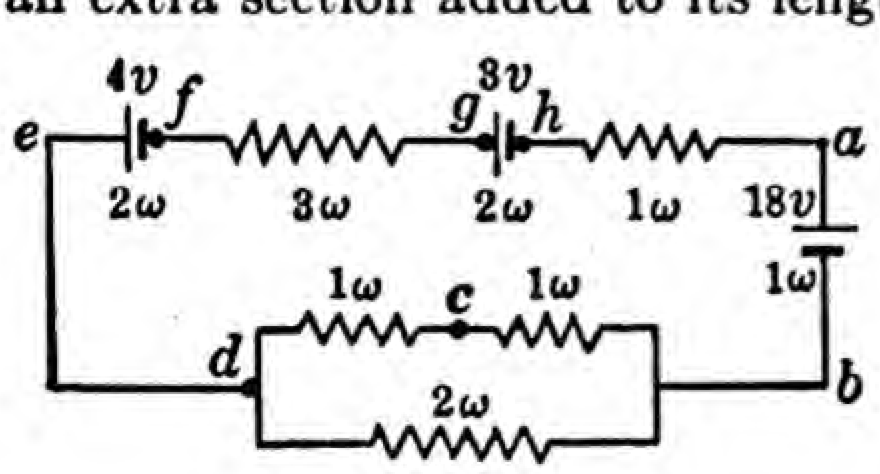


FIG. 130

5. In Fig. 130, find the potential difference across the following points and tell which point has the higher potential:

ab, bd, cd, ef, eh, fa.

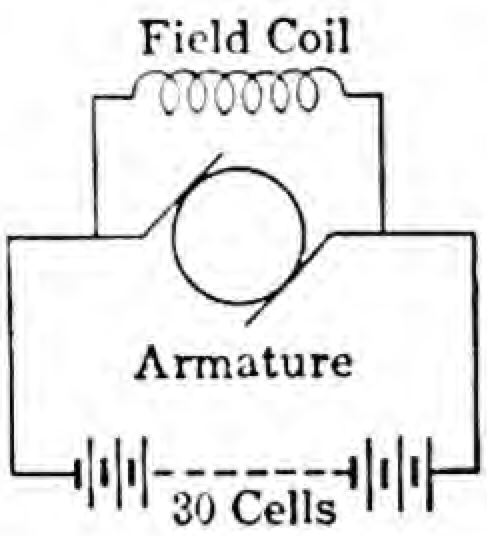


FIG. 131

6. A d.c. generator whose field coil and armature coil are in parallel has an e.m.f. of 100 volts and an armature resistance of 2 ohms (Fig. 131). It is used to charge a storage battery of 30 cells, each having an e.m.f. of 2 volts and an internal resistance of 0.1 ohm. The field coil has a resistance of 30 ohms. Compute the current flowing through the battery.

7. Compute the magnitude and direction of the current in each branch of the following circuits (Fig. 132). In part *d* compute by two different paths the difference in potential between the upper right and lower left-hand corners of the figures.

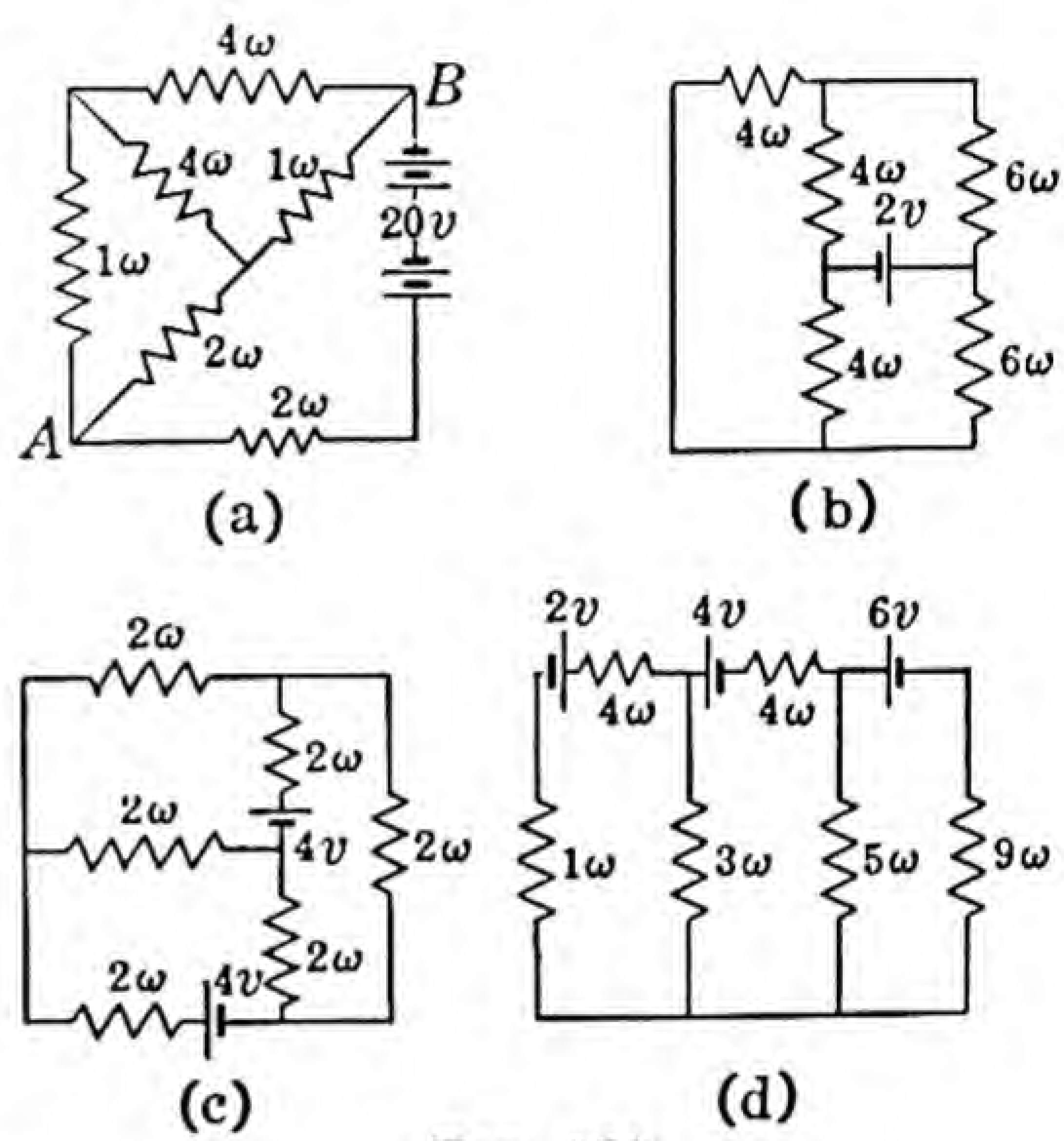


FIG. 132

8. Find the effective resistance between the points *A* and *B* in part (*a*) of Problem 7.

9. Twelve straight wires of 3 ohms each are connected so as to outline a cube. What is the resistance between opposite corners?

10. Is the circuit shown in Fig. 133 solvable without Kirchhoff's laws? Find the current in each branch.

11. A power house is located 1000 ft. from a building in which there are 500 60-watt, 115 volt lamps. If it is specified that only 5 per cent of the energy delivered to the line shall be lost in the line and that the voltage at the building be full 115 volts, what must be the resistance of the line and the diameter of the copper wire used? If the resistance of the generator is 0.01ω, what must be the e.m.f. generated, the copper loss (watts dissipated as heat in the resistance of the copper windings) in the generator, and the potential difference across its terminals?

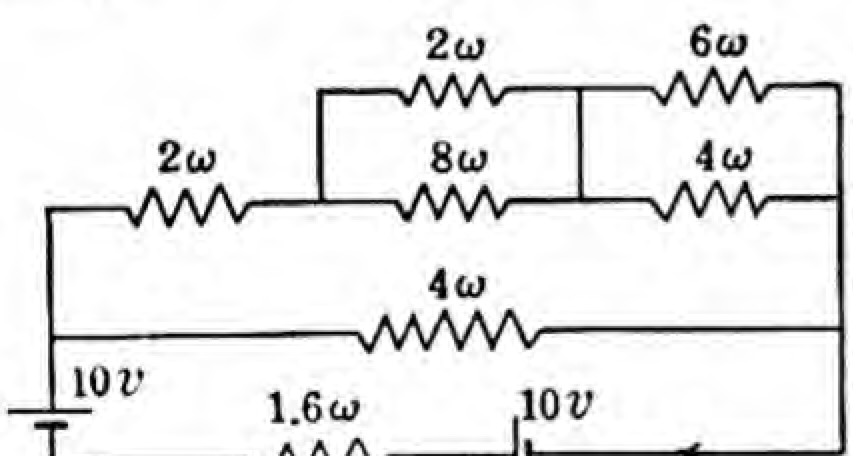


FIG. 133

12. Point *C* (Fig. 134) is 14 volts higher than *D*. Find V_{AB}, V_{BC}, and the value of E_2. How many calories of heat will be generated in part *AB* in 10 minutes?

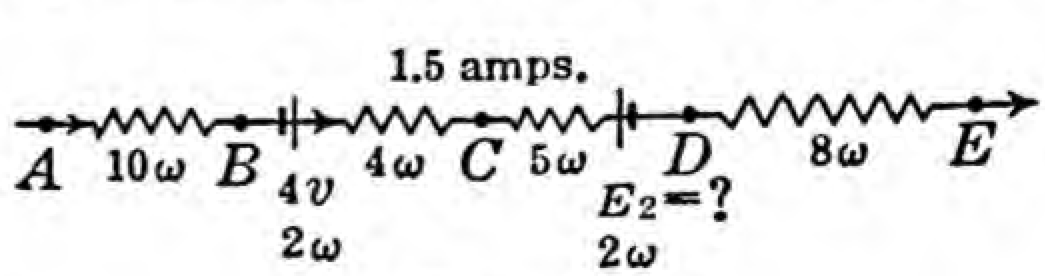


FIG. 134

13. An electric water heater (a suitable coil of wire) is immersed in a vessel containing 800 gm. of water at 20° C., and a potential difference of 110 volts is applied. The water begins to boil after 5 minutes. Neglecting the heat capacity of calorimeter and the loss due to radiation, find (*a*) the power applied; (*b*) the resistance of coil.

14. Fig. 135 shows a single line circuit containing a generator, G, and lamps, L, the return being through the ground. The potential difference across the terminals of the generator is 110 volts. Each lamp takes 2 amperes. The distance from A to B is 2000 ft., from B to C is 1000 ft., and from C to D 3000 ft. The line wire is of copper 200 mils in diameter. Find the potential at D. Consider the ground resistance for each distance to be equal to that of the corresponding length of copper wire.

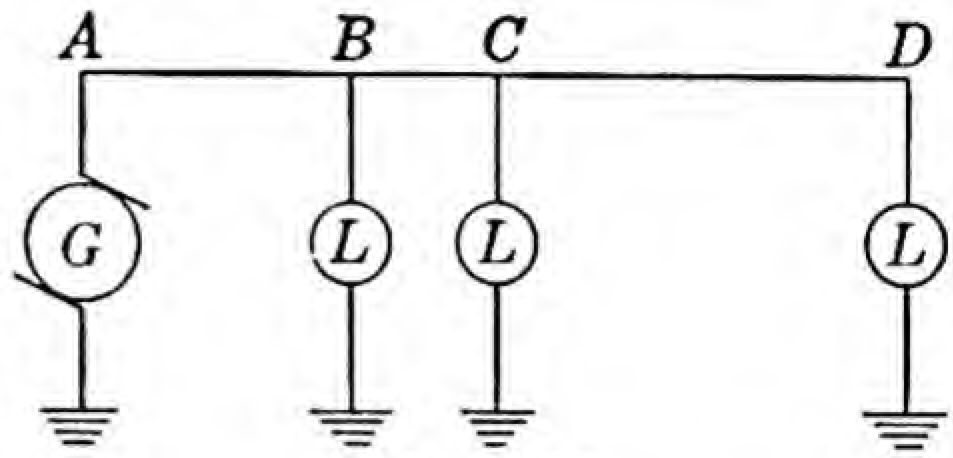


FIG. 135

15. A platinum wire used as a resistance thermometer has a resistance of 10 ohms at 0° C.; when placed in a furnace the resistance is 28.5 ohms. The temperature coefficient of the resistance of platinum is 0.0039 per ° C. What is the temperature of the furnace?

ELECTRICAL MEASURING INSTRUMENTS

155. The Tangent Galvanometer. — The tangent galvanometer consists of a compass needle mounted at the center of a vertical circular coil of wire. The coil is placed so that its plane is parallel to the earth's field. Then the current which is to be measured is sent through the coil and the magnetic field produced at the center of the coil is given by Eq. 213. This field is at right angles to the earth's field, so we have the case discussed in § 131. The compass needle deflects through an angle θ and combining Eqs. 199 and 213 we get

$$\frac{H_A}{H_e} = \frac{\frac{2\pi Ni}{r}}{H_e} = \tan\theta,$$

or

$$i = \frac{rH_e}{2\pi N}\tan\theta,$$

$$I = \frac{10\,rH_e}{2\pi N}\tan\theta. \tag{238}$$

So if the radius of the coil r is measured and the value of the earth's field is measured, the strength of the current is determined. The instrument receives its name from the form of the equation.

After the current is turned on, it is possible to turn the plane of the coil in the direction of the deflection of the needle until the coil is again in line with the needle. The resultant of the two fields then has the direction of the plane of the coil, and the student should prove that the current is proportional to the sine of the angle of rotation of the coil (or needle). The instrument

designed with a divided circular scale to measure the angular displacement of the coil is called the Sine Galvanometer.

It is seen that the tangent or sine galvanometer gives a means of determining by an absolute method the strength of an unknown current. We have seen in § 132 how H_e may be determined in absolute measure, and from Eq. 238 it is seen that, in order to determine I absolutely, the only other quantities necessary to be measured are the radius of the coil (length), the number of turns of wire (no dimensions), and an angle θ (ratio of two lengths).

Now after I is known, an absolute measurement of resistance may be made by measuring the heat developed in a given time by a known current according to Eq. 220. The time t is measured in seconds, and the energy is measured in calories by a calorimeter. Calorimetric measurements are made entirely by measurements of mass and distances along a thermometer scale. Heat measurements are not capable of as great precision as electrical measurements but there are other absolute methods for measuring resistance that give highly accurate determinations.

As soon as I and R are determined then all other electrical units are determinable, V from Eq. 221, etc.

156. Torque on a Plane Coil Suspended in a Uniform Magnetic Field. — Consider a uniform magnetic field of strength H, Fig. 136, parallel to the xz plane but making an angle δ with a plane coil placed in the xy plane. We wish to determine the torque on the coil about the y axis.

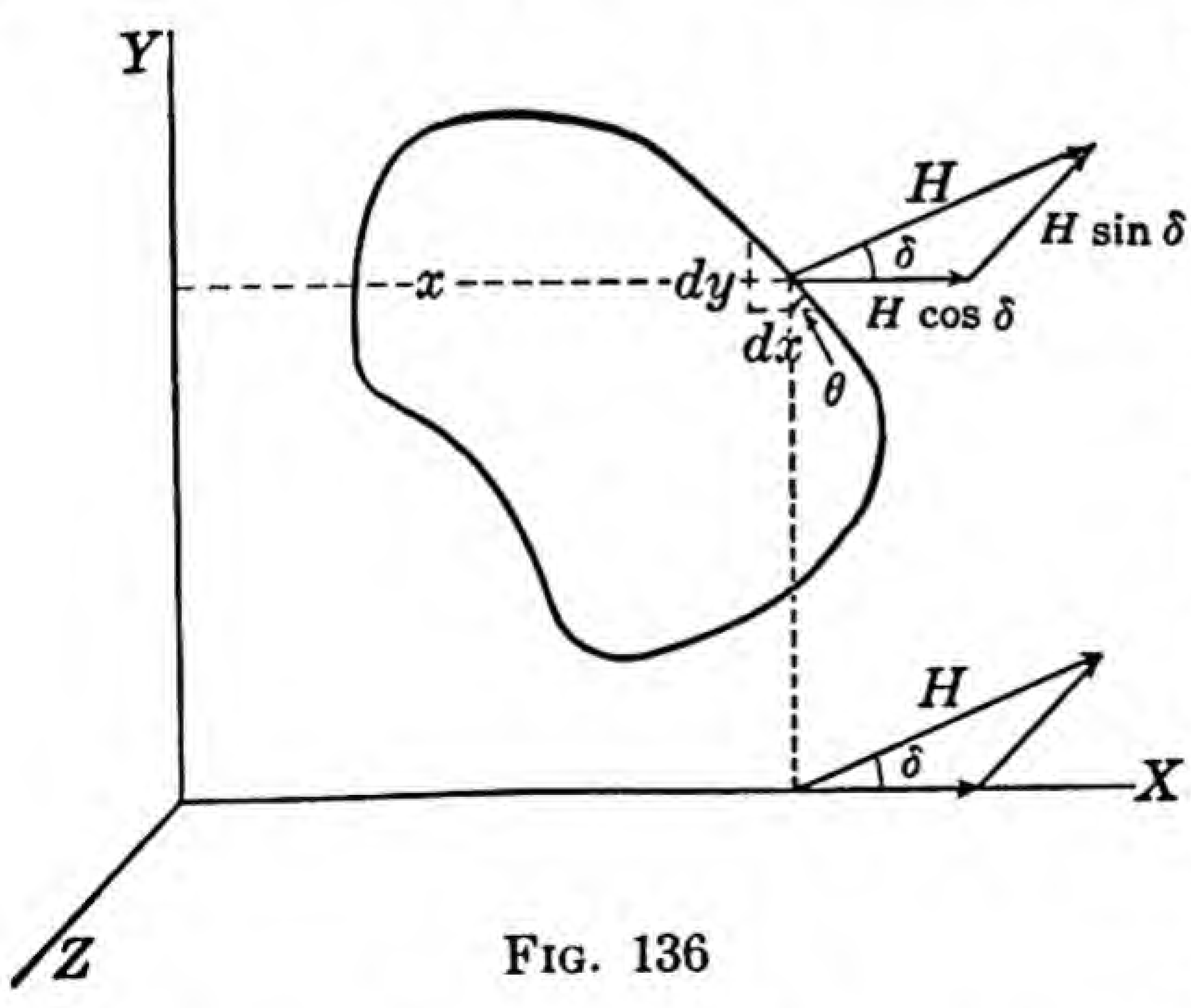


FIG. 136

First let us resolve the field into its two components, $H_z = H \sin \delta$ and $H_x = H \cos \delta$. The component H_z acts on each element of current only so as to tend to stretch or shrink the coil in its own plane and so cannot produce any torque about the y axis. The component H_x, however, will produce a force perpendicular to the plane of the coil and hence will produce a torque about the y axis. Consider a section of the coil of length

ds having components dx and dy and abscissa x. According to Eq. 211 the force on the element ds is $H_x i \, ds \sin \theta$. The moment of this force about the y axis is

$$dL = H_x ix \, ds \sin \theta$$
$$= H_x ix \, dy. \tag{239}$$

Adding up the moments for all elements around the whole closed loop of the coil, we get

$$L = iH_x \int x \, dy. \tag{240}$$

The expression $\int x \, dy$ gives the area A of the coil. For N turns of wire on the coil, the torque is

$$L = NiAH_x = \frac{NIAH}{10} \cos \delta. \tag{241}$$

157. The d'Arsonval Galvanometer. — A d'Arsonval galvanometer consists merely of a coil of wire suspended between strong permanent magnets. The usual design is as shown in Fig. 137. The pole pieces of the permanent magnets are shaped so that the coil will just turn freely without touching them. A soft-iron cylinder is supported inside the coil, producing a field in the air gap such that if the lines of force were extended without bending (as they do in the iron), they would all cross at the center of the cylinder. The field is said to be radial and in such a field the plane of the coil makes at all times a zero angle with the field. Then in Eq. 241, δ will be zero. The coil is usually suspended by a fine metallic ribbon through which the current enters the suspended coil. The current flows out through a fine helix of the same ribbon at the bottom of the coil. As the coil turns under the action of the torque produced by the current, the suspension twists and the coil comes to a position of equilibrium when the twisted suspension exerts a torque on the coil equal and opposite to that due to the current. If L_0 (see § 54) is the torsion constant of the suspension and α the angle of deflection of the coil, $L_0\alpha$ is the torque due to the suspension. Then

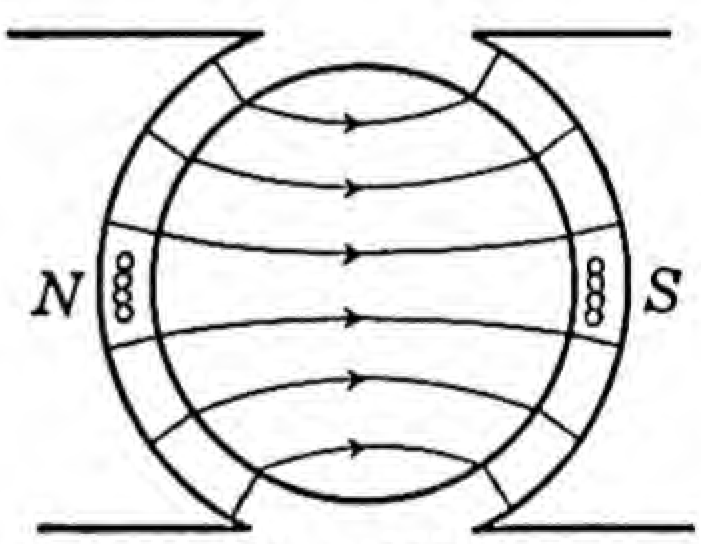


FIG. 137

$$L = L_0\alpha = NiAH,$$

or

$$i = \frac{I}{10} = \frac{L_0}{NAH}\alpha = k\alpha. \tag{242}$$

So for a d'Arsonval galvanometer constructed so as to have a radial field, the current is directly proportional to the deflection of the coil.

158. The Wheatstone Bridge. — Four resistances, a galvanometer, and a battery connected as in Fig. 138 form a Wheatstone Bridge. The battery maintains a potential difference between the points a and b. Since there is the drop in potential along each branch of the circuit, it is possible to find points c and d whose potentials are the same and then no current can flow through the galvanometer. If R is then decreased, the potential at d is less than that at c (why?) and current will flow down from c through the galvanometer. If R is increased, the current will flow up from d through the galvanometer. The bridge is said to be "balanced" when there is no current through the galvanometer. In this case only, the same current flows through R_1 and x and likewise for R_2 and R. For the balanced condition,

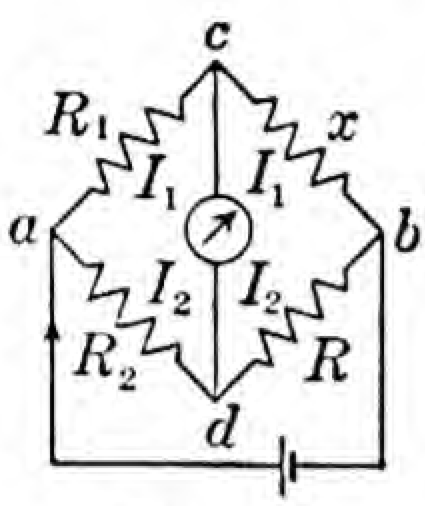


FIG. 138

$$I_1R_1 = I_2R_2,$$

and

$$I_1x = I_2R.$$

Hence

$$\frac{R_1}{R_2} = \frac{x}{R},$$

or

$$x = \frac{R_1}{R_2}R. \tag{243}$$

The Wheatstone Bridge is thus a device for comparing four resistances. In order to find the value of x either all the other three resistances must be known or the ratio of R_1 to R_2 and the value of R must be known.

The student should interchange the battery and galvanometer and show that the same relations hold.

There are many networks used in measuring electrical quantities in which, after the current in the galvanometer has been reduced to zero, the battery and galvanometer may be interchanged and the circuit is still balanced. Any such arrangement is called a "bridge."

159. The Potentiometer. — The principle of the potentiometer is shown in Fig. 139. The main circuit consists of a battery E and a control rheostat to vary the current I through the main resistance ab. Two cells of e.m.f. E_s and E_x are connected as

shown. When the switch S is closed upward, the standard cell E_s is connected across the points a and c of the main resistance. The battery E tends to send current to the right through E_s, and E_s tends to send current to the left through the resistance from a to c. If the point c is moved until the potential drop $V_{ac} = E_s$, these opposing effects balance each other and then no current flows from E_s and the galvanometer indicates zero deflection. For all points on the two opposite sides of c, the galvanometer shows opposite deflections. When the switch S is closed downward, then some point c' may be found where the galvanometer is not deflected and $V_{ac}' = E_x$.

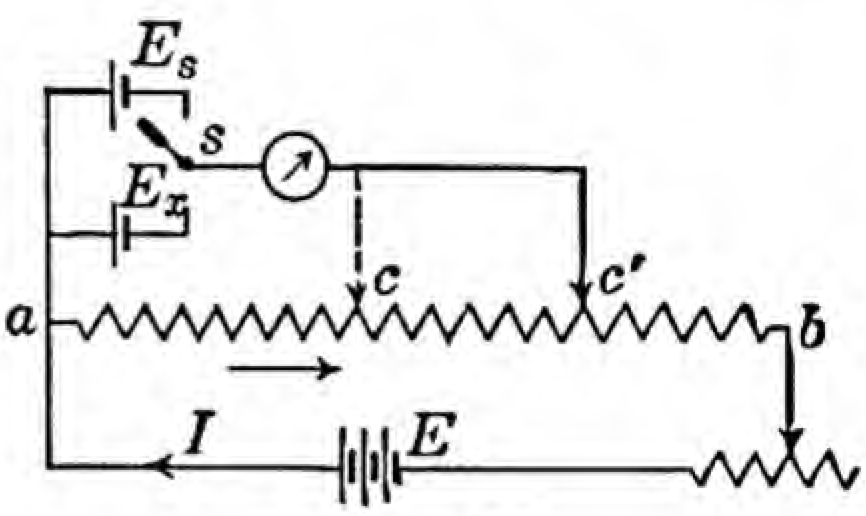


FIG. 139

$$\frac{E_s}{E_x} = \frac{IR_{ac}}{IR_{ac'}} = \frac{R_{ac}}{R_{ac'}}. \tag{244}$$

Obviously E must be larger than E_s and E_x each, in order that balances be obtainable.

The student should apply Kirchhoff's laws to the circuit and obtain Eq. 244.

The potentiometer may be used to compare one e.m.f. with another. It is essential that the current I remain constant during the two processes of "balancing" the potentiometer or otherwise the two currents will not cancel in the above equation and then comparison of the e.m.f. cannot be made. If one of these is a standard cell measured by the Bureau of Standards and certified to be 1.01875 volts, say, then the unknown cell can be measured to the same degree of accuracy as that of the standard, provided the ratio of the resistances in the potentiometer is known to that same percentage of accuracy.

The main advantage of the potentiometer is due to the fact that when the final settings are made no current is being drawn from the cells which are being compared. In § 149, we saw that when current is being drawn from a cell, the potential difference across its terminals is less than its e.m.f. Hence the potentiometer compares the actual e.m.f.'s of the cells.

160. Ammeters and Voltmeters. — The main portion of most direct-current ammeters or voltmeters is a d'Arsonval galvanometer. The moving coil is supported on jeweled bearings. The

current is led into and out of the coil through spiral wires and these wires produce the necessary return torque. The arrangement of an ammeter is shown inside the dotted circle in Fig. 140.

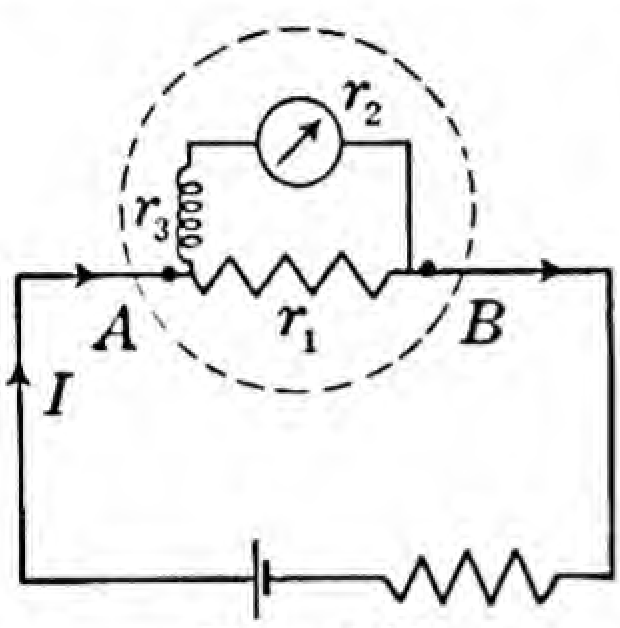


FIG. 140

The galvanometer is shunted with a very low resistance r_1. The ammeter is connected by its binding posts A and B so as to be in series with the main circuit in which the strength of the current is to be measured. An ammeter must always have as low a resistance as possible in order that it will not increase the resistance of a circuit into which it is inserted and thus decrease the current to be measured. Most of the current passes through the shunt r_1. The potential drop across this shunt resistance r_1, although small, causes a small current to flow through the galvanometer and the pointer fastened to the coil registers a reading on the scale. The fine wire wound on the moving coil has a high resistance r_2 and then in addition there is a spool of resistance r_3 connected in series with the galvanometer. The amount of the resistance of the wire on the spools is varied by the manufacturer in order to make up for slight differences in the sensitivity of the galvanometers and thus all the instruments of one lot are made to fit the stamped scales.

A voltmeter, on the other hand, always has a high resistance and is placed in parallel across the points of a circuit across which the potential difference is to be measured, Fig. 141. If the voltmeter were of low resistance, it would lower the effective resistance between the points a and b and hence change the current I through the main circuit and change the potential difference whose measurement was desired. So the voltmeter consists of a galvanometer in series with a very high resistance R. The resistance of the voltmeter (from A to B) should always be many times larger than the resistance of the circuit, such as ab, across which it is placed. Although the voltmeter instrument has a high resistance it must have enough current flowing directly through it to deflect the coil of the galvanometer.

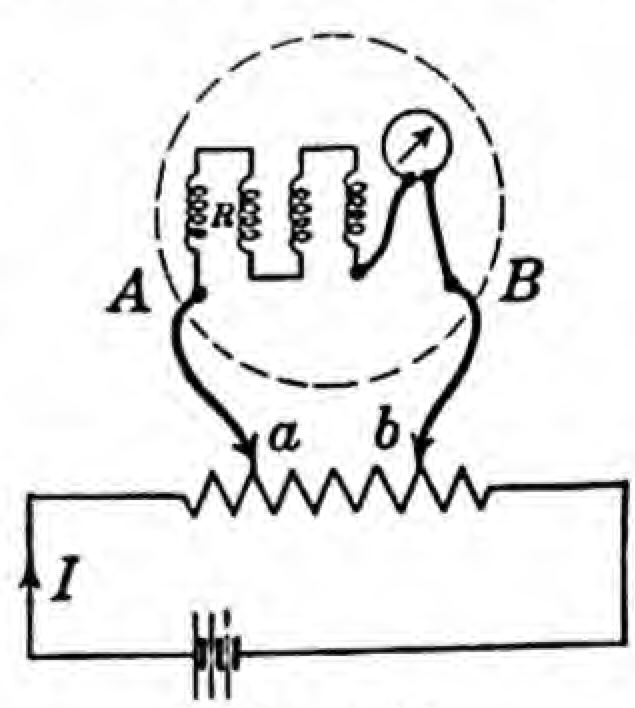


FIG. 141

The range of a voltmeter is easily extended. If it is desired to read voltages up to n times the largest value on the scale of a

voltmeter, it is only necessary to put in series with the voltmeter a resistance $(n - 1)$ times that of the meter itself. Such a resistance, adjusted to match a given voltmeter, is appropriately called a *multiplier*.

A millivoltmeter is a very sensitive galvanometer with a small resistance in series with it. It must give a full-scale deflection when placed across two points of a circuit whose potential difference is only a certain number of thousandths of a volt. Such an instrument may be used with a set of shunts as an ammeter of various ranges. Thus if the instrument has a scale reading up to 100 millivolts, then a shunt of 0.001 ohm would make the instrument into an ammeter reading up to 100 amperes because 100 amperes $\times$ 0.001 ohm = 0.100 volt = 100 millivolts and that potential would cause the full-scale deflection of the millivoltmeter.

A very high resistance voltmeter must have a very sensitive galvanometer and so such an instrument might be used as an ammeter for small currents when the resistance is not objectionable.

PROBLEMS

1. All shunts, such as just described for converting millivoltmeters into ammeters, have four binding posts. Two are used to lead the current in and out and the other two for wires leading to the millivoltmeter. Why would not two posts be sufficient? Why must a millivoltmeter have a given special set of wires supplied with it for use in connecting the meter to the shunts?

2. A current is sent through a standard resistance and an unknown resistance in series. How could a potentiometer be used to compare the two resistances?

3. Suppose that a 0–150 volt voltmeter whose resistance is 15,000 ohms is to be used to measure a potential difference of 600 volts. Let a subdivided megohm resistor be placed across the 600 volt line and the voltmeter be connected across one quarter of the megohm. Will the voltmeter read 150 volts and thus indicate 4 $\times$ 150 or 600 volts? What will be its reading?

4. A current of 0.01 amp. causes a deflection of 20 divisions on the scale of a milliammeter. The instrument has a resistance of 50 ohms and there are 100 divisions on the scale. What resistance in series with the meter will enable it to measure up to 100 volts?

5. A 0–50 milliammeter has a resistance of 10 ohms. What shunt resistance will convert the meter into a 0–50 ammeter?

6. The scale of a certain d'Arsonval galvanometer is 25 cm. long. The galvanometer has a resistance of 100 ohms and a sensitivity of 8×10^{-5} amperes per cm. What resistances, and how connected, will enable the instrument to be used as a triple range voltmeter with 2.5, 25, and 125 volt ranges?

7. In Problem 6, it is desired to measure voltages up to 40 volts with the greatest accuracy allowed by the instrument. What external resistance must be put in series with the 25 volt tap for this use?

8. It is desired to use the galvanometer in Problem 6 as a direct-reading milliameter (0–25 m.a.), millivoltmeter, ammeter, and voltmeter. Are all of these cases possible? What arrangements of resistances are necessary?

9. Suppose that the galvanometer in Problem 6 and a standard 0.01 ohm resistance are available. In what two ways may a third resistance be connected in order to make a 0–25 ammeter? Find the two values of the third resistance. Which is the more practical means?

10. A good standard voltmeter usually has a resistance of at least 100 ohms for each volt of its maximum scale reading. What is the least current required for a full-scale deflection of a galvanometer which may be used for such instruments? If a voltmeter reading up to 150 volts has exactly 100 ohms per volt, what current flows through it when it reads 15 volts?

11. For measuring the voltage of B battery "eliminators" a voltmeter of 1000 ohms per volt should be used. Why? What should be the resistance of such a voltmeter designed to read up to 500 volts? What is the least sensitivity (amperes per scale division) for the galvanometer whose scale has 100 divisions?

12. A certain standard cell, $E = 1.01870$ volts, has 1000 ohms internal resistance. No standard cell should be used where it will supply more than 10^{-5} amperes. A voltmeter of the type in Problem 11 has a range of 3 volts. (*a*) If it were connected to the standard cell, what voltage would it read? (*b*) Would the standard cell be injured? (*c*) What minimum resistance must a 3 volt voltmeter have and what would the minimum sensitivity of the galvanometer unit have to be in order that it might not damage the cell? (150 divisions on scale.) (*d*) Under these conditions what would the voltmeter reading be and what would be the per cent error?

Induced Electromotive Forces

161. Lenz's Law. — We owe to Faraday the first experimental demonstration that an electric current will flow in a closed circuit whenever any portion of such a circuit is moved across a magnetic field or whenever a magnetic field in the neighborhood of such a circuit is caused to change. Such currents are called "induced currents" and the corresponding electromotive forces are called "induced electromotive forces." A very important generalization on the laws of electromagnetic induction made by Lenz states that *the induced current always flows in a direction so as to oppose, by its electromagnetic action, whatever causes the current to be induced.*

Consider a conductor which is part of a closed circuit and which is placed at right angles to a magnetic field. When the conductor is pushed so as to cut across the lines of force, a current flows

through it in a direction so that, in front of it, the field due to the current is in the same direction as the original field, and behind it is in a direction opposite to that of the original field. Thus a strong field exists in front of the wire and a weak field behind it, and we have seen that such a condition causes a force directed toward the weaker field, or opposing the motion. Therefore work has to be done in order to supply the energy represented by the electric current. If there were not this opposition or if the force were so as to help the motion, then the conductor once started would, in the first case, continue in motion creating electrical energy continuously or, in the second case, doing mechanical work continuously in addition to supplying the electrical energy. Thus Lenz's law is merely the statement of the principle of conservation of energy as applied to electromagnetic induction.

162. The Magnitude of the Induced E.M.F. — Let a straight portion of a closed conducting circuit of length l be moving with velocity $v = ds/dt$ at right angles to a field of strength H. This may be done practically by having two wires, AB and CD (Fig. 142), of negligible resistance placed a distance l apart. A short-circuiting bar is slid along these rods with a velocity v. The induced e.m.f., e, causes a current i to flow. In the time dt an amount of energy $ei\ dt$ is produced by mechanical work done against the opposing electromagnetic force. This force by Eq. 211 is μHil. In order to move the conductor with a constant velocity v a force F, equal and opposite to this force, must be applied. Then the work done during an infinitesimal displacement becomes

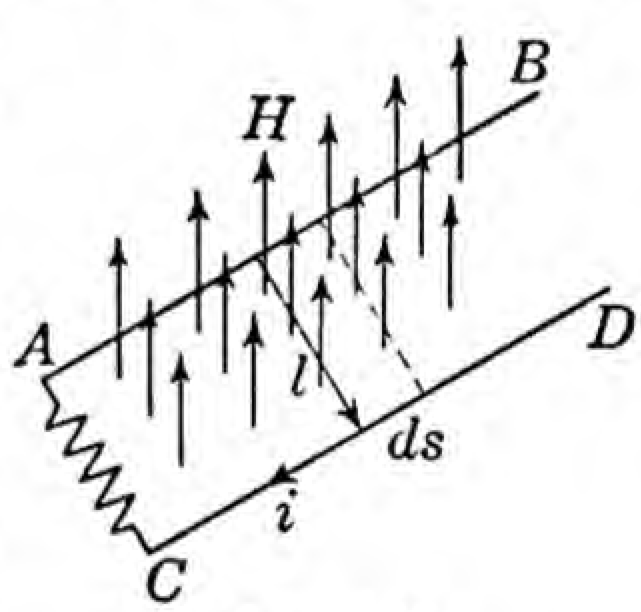


FIG. 142

$$dW = F\,ds = \mu Hil\,ds = ei\,dt. \tag{245}$$

Hence,
$$e = \mu Hl\frac{ds}{dt} = \mu Hlv. \tag{246}$$

Now $l\,ds$ represents the area swept over by the conductor of one turn and $\mu Hl\,ds$ represents the total number of lines, $d\Phi$, cut during the time dt. So

$$e = \mu Hlv = \frac{d\Phi}{dt} \text{ for one turn,}$$

or
$$e = N\mu Hlv = \frac{N\,d\Phi}{dt} \text{ for } N \text{ turns.} \tag{247}$$

In Eq. 245, if H is expressed in gauss, i in e.m.u., l and ds in cm., then $F = \mu Hil$ is in dynes and so dW is in ergs. In order for the right-hand member to give a result in ergs, both i and e must be expressed in e.m.u. Hence in Eq. 247 e is in e.m.u. and we see, therefore, that *when one line of force is cut per second, there is induced an electromotive force of one e.m.u.*

Knowing from § 141 that 10^8 e.m.u. = 1 volt, we have

$$E \text{ volts} = \frac{N\,d\Phi}{10^8\,dt}. \tag{248}$$

where $d\Phi$ represents the change in flux through each turn.

163. Quantity Induced. — Consider a circuit which enclosed flux Φ_1 and which after a certain time enclosed flux Φ_2. During the change there must have been an induced e.m.f. proportional at all times to the rate of change of flux through the circuit. If the circuit is closed and has a resistance r, then a certain current flows and we may obtain the total quantity of electricity transferred as follows:

$$q = \int dq = \int_0^t i\,dt = \int_0^t \frac{e}{r}\,dt = \int_{\Phi_1}^{\Phi_2} \frac{N\,d\Phi}{r} = \frac{N(\Phi_2 - \Phi_1)}{r}. \tag{249}$$

The quantity q is expressed in e.m.u. if also are i, e, and r. If q is desired in coulombs, i must be expressed in amperes, e in volts, and r in ohms, and then, using capital letters for practical units,

$$Q = \frac{N(\Phi_2 - \Phi_1)}{10^8\,R}. \tag{250}$$

The quantity $N\Phi$ is called the *flux linkage.*

164. The Alternating Current Generator. — *The E.M.F. in a Plane Coil of Any Shape Rotating about an Axis Perpendicular to a Uniform Magnetic Field.*

Let a plane coil of any shape, of area A, be rotating about any axis perpendicular to the direction of the field (Fig. 143). Let $N\Phi_m$ be the maximum flux linkage possible, when the coil is perpendicular to the magnetic field. When the coil makes an angle θ with a plane perpendicular to the field, the flux linkage is $N\Phi_m \cos\theta$ and then by Eq. 248,

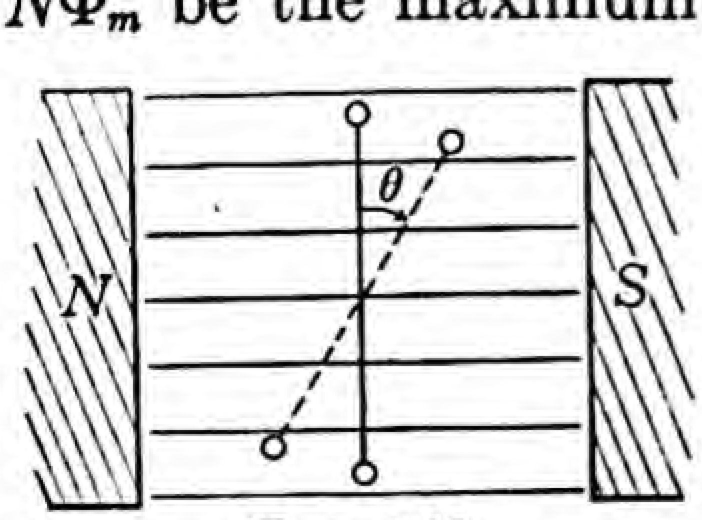


FIG. 143

$$E = \frac{d(N\Phi_m \cos\theta)}{10^8\,dt} = \frac{N\Phi_m \sin\theta\,d\theta}{10^8\,dt}.$$

Now $\theta = \omega t = 2\pi nt$, where ω is the angular velocity of the coil and n is the number of revolutions per second. Then

$$E = \frac{2\pi nN\Phi_m}{10^8}\sin 2\pi nt = E_0 \sin\theta = E_0 \sin \omega t, \tag{251}$$

where E_0 is the maximum value of the e.m.f. which occurs at the instant when the coil is in the plane of the magnetic field so that the conductors in which the e.m.f. is induced are moving at right angles to the field. The e.m.f. is alternating and Eq. 251 is the fundamental equation of an alternating current generator.

If the ends of the coil A are connected to slip rings as shown in Fig. 144 and connection is made to an external circuit from the rings, an e.m.f. $E_0 \sin \omega t$ exists in the circuit and, after transient effects have died out, a current $I = I_0 \sin \omega t$ will flow. The maximum value of I is I_0. The current varies as shown by the curve A in Fig. 145. It increases from zero to I_0, decreases to zero, increases to a maximum in the opposite direction, etc. The generator pictured in Figs. 143 and 144 has two poles, one north pole and one south pole. It is called a two-pole single-phase generator.

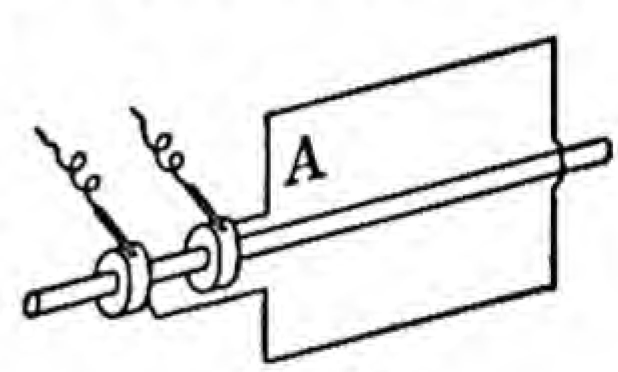


Fig. 144

If another coil, B, is mounted on the same armature but placed at right angles to coil A and connected to a separate pair of slip rings, we have a two-pole two-phase generator. In practice only three slip rings are used, one of them being used in common by the two circuits. The currents in the two external circuits are related as shown in Fig. 145. When I_A is maximum, $I_B = 0$. As I_B reaches its maximum, one quarter of a revolution of the armature later, I_A has decreased to zero, and so on. They are said to be 90° apart in phase. Of course the coils are 90° apart geometrically but this is the case only for a two-pole generator. If there were four poles alternately north and south, there would be two complete alternations of the current per revolution and the planes of the two coils would have to be 45° apart in order to have their currents 90° out of phase, as represented in Fig. 145. With three separate coils placed 120° apart for a two-pole machine, or 60°

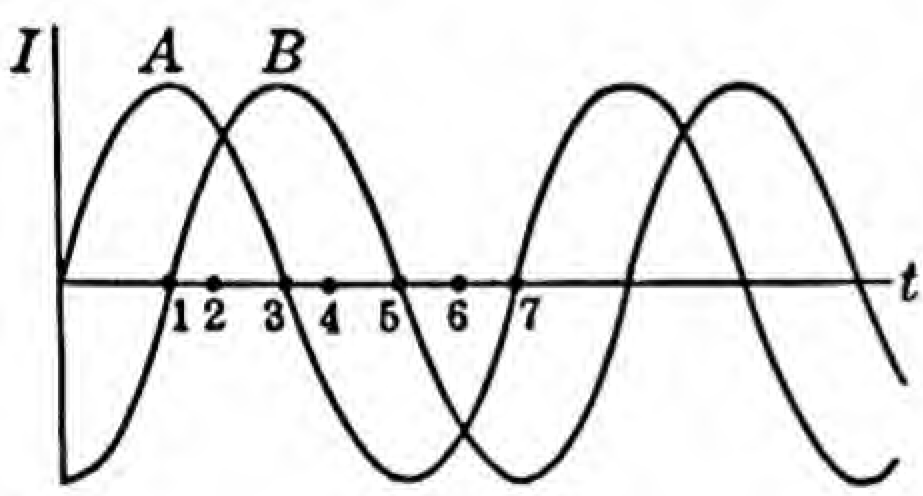


Fig. 145

apart for a four-pole machine, three-phase current may be obtained with the currents in adjacent coils 120° out of phase. Further discussion of polyphase multipolar machines belongs to the field of electrical engineering.

165. The Direct-Current Motor. — When a current flows through a loop of the armature placed in a radial magnetic field, the loop is acted upon by a torque, which may be computed from Eq. 241. If the armature rotates under the action of this torque, every loop then cuts across the field and an e.m.f., E_m, is induced. By Lenz's law this e.m.f. must be in such a direction as to oppose the motion, and so it must be opposed to the potential difference Pd_m, applied to the terminals of the motor, which causes the current to flow. As the motor starts from rest and picks up speed, the opposing electromotive force E_m increases and the current is decreased. This process continues until equilibrium is reached.

The amount of power developed by the motor may be determined by a proof which is somewhat the reverse of that in § 162. There we saw that the mechanical motion generated an e.m.f. which caused a current i to flow, delivering power ei. The e.m.f. was directed so that the current flowed in a direction to produce forces opposing the motion. In the reverse case, we can reason that when the current i flows, forces exist which cause motion. The motion causes an e.m.f. opposing the flow of current. Since we know that mechanical work can be done only in case of motion against opposing forces, we see that the power developed by the motor must be $e_m i$. We may incorporate this logic in the following proof, where l is the total length of wire that cuts the field.

$$\text{Power} = \text{Force} \times \text{velocity} = Hilv.$$

By Eq. 246, Hlv is the generated e.m.f. which opposes the current flow. So

$$\text{Power} = e_m i \text{ ergs/sec.} = E_m I \text{ joules/sec. or watts.}$$

166. Efficiency of a Series-Wound Direct-Current Motor. — The efficiency of a motor is defined to be the ratio of the amount of work done by the motor to the amount of electrical energy supplied to the motor.

$$\text{Electrical Efficiency} = \frac{\text{Output}}{\text{Input}}$$

$$= \frac{E_m I}{Pd_m I} = \frac{E_m}{Pd_m}. \qquad (252)$$

When the armature of the motor is held still,

$$E_m = 0, \quad \text{and} \quad I = \frac{Pd_m}{R_m},$$

a very large current. The output is zero and the efficiency is zero. When there is practically no load on the motor, the armature speed increases to a maximum E_m approaching Pd_m and I approaching zero. In this latter case, the efficiency is seen by Eq. 252 to be nearly 100 per cent. This high efficiency, however, is of little value, because under the circumstances, there is no output. It is obvious that if the output is zero at both zero and maximum speeds, at some point in between there will occur a maximum rate of doing work. The speed at which the maximum output is obtained could be found by equating to zero the derivative of E_mI with respect to the speed of the armature. But since E_m is directly proportional to the speed of the armature we may find the condition by writing

$$\frac{d(E_mI)}{dE_m} = 0.$$

The differentiation cannot be carried out at once because the value of I depends on the value of E_m. Then we must first express I as a function of E_m. Neglecting iron losses and friction losses, $I = (Pd_m - E_m)/R_m$. Therefore

$$\text{Output} = E_mI = E_m\left(\frac{Pd_m - E_m}{R_m}\right) = \frac{Pd_mE_m - E_m^2}{R_m}.$$

The student should show, as outlined above, that the output is a maximum when $E_m = Pd_m/2$ or when the efficiency is 50 per cent. What practical reasons would prevent large-size motors from being thus operated, apart from the cost?

167. The Induction Motor. — We have seen that when a conductor which is part of a closed circuit is moved across a magnetic field, a current is induced so as to oppose the motion. The same conditions apply if the conductor is held still and the field is moved. If free to move, the conductor will follow the field. If it is necessary to do work in moving the conductor, then the conductor will lag behind the field so that lines of induction will be cut and an e.m.f., E_m, will be induced. The e.m.f. will cause a current I to flow. The conductor will lag sufficiently that E_mI will equal the power required to move the conductor.

A practical application of this principle is the alternating-current induction motor in which the field is caused to rotate continuously and the armature consists of a "squirrel cage" arrangement of conductors perpendicular to the field. A four-pole two-phase motor of this type is illustrated in Fig. 146. The alternating current through the coils AA' must be 90° out of phase with the current in the coils BB' (see Fig. 145). Let the phase of the current in the B coils be 90° behind that in the A coils. When the current in A is such that the field at P is a maximum to the right, the current (and hence its field at P due to it) is zero through B (Point 1 in Fig. 145). As I_A decreases, the current through B starts increasing (Point 2 in Fig. 145). Let the connections at B and B' be such that when the field due to A has decreased to zero, the field due to B has reached its maximum upward value (Point 3 in Fig. 145). Then the field due to A starts increasing toward the left as that due to B decreases, etc. We shall now prove that the magnitude of the resultant of these two fields is constant when the maximum value of the current is the same in both coils.

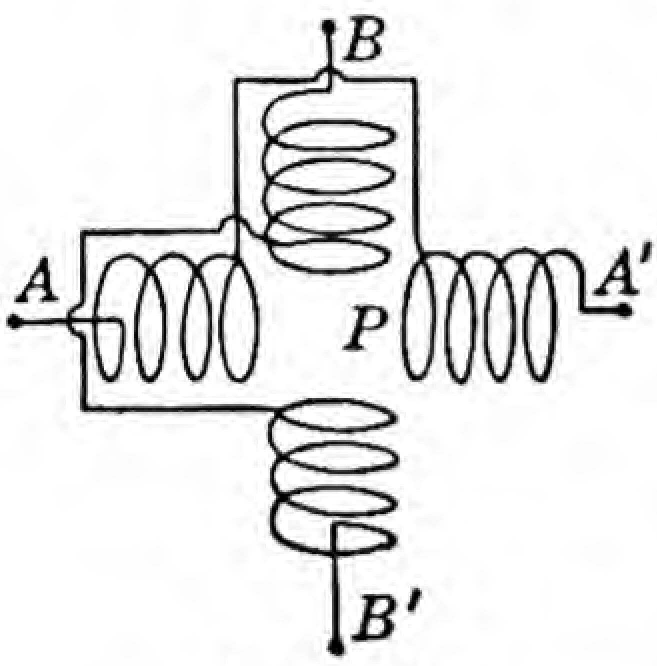


FIG. 146

The current in coil A may be expressed by Eq. 251 as

$$I_A = I_0 \sin \omega t.$$

For the current in B,

$$I_B = I_0 \sin\left(\omega t - \frac{\pi}{2}\right) = - I_0 \cos \omega t.$$

Let H be the field produced by the current I and let its maximum value be H_0 when $I = I_0$. Then

$$H_A = H_0 \sin \omega t, \text{ and } H_B = -H_0 \cos \omega t.$$

So $$H = \sqrt{H_A^2 + H_B^2}$$
$$= \sqrt{H_0^2 (\sin^2 \omega t + \cos^2 \omega t)} = H_0,$$

which is constant. So the effect of the two alternating fields is to produce a rotating magnetic field of constant strength. The relation between the individual fields and their resultant is shown in Fig. 147. The numbers around the circles and the subscripts correspond to the numbers in Fig. 145.

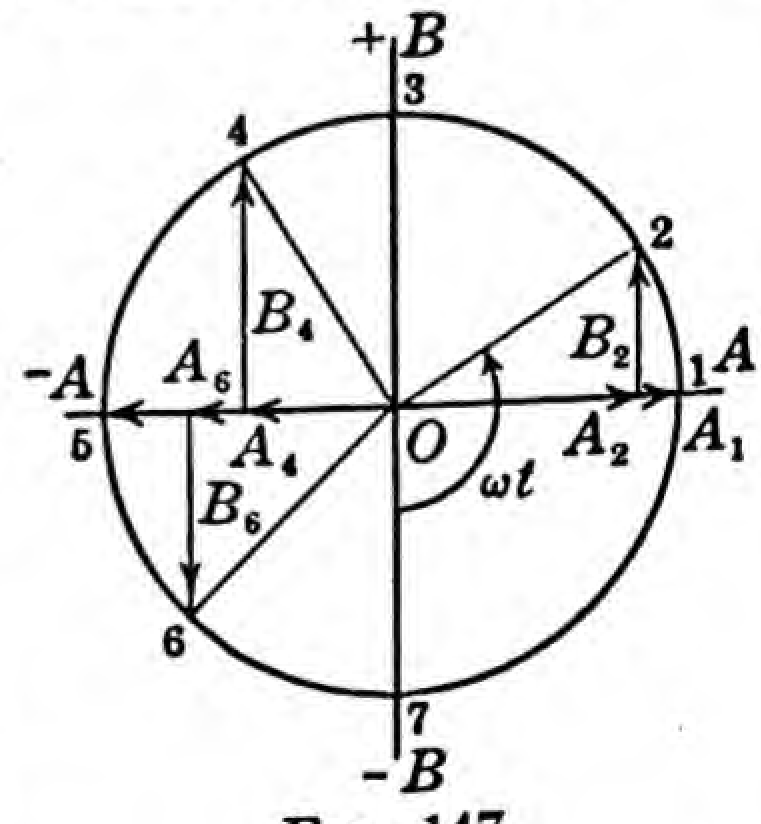


FIG. 147

PROBLEMS

1. One coil of a dynamo armature such as in Fig. 144 has a length of 40 cm., a width of 30 cm., and has 20 turns. It rotates at 1500 r.p.m. in a magnetic field of 1000 gauss. Find the maximum e.m.f. generated in the coil.

2. A solid copper disk 1 meter in diameter rotates on a horizontal axle which is directed toward the geographic north at a place where the horizontal intensity of the earth's field is 0.18 gauss. It makes 250 r.p.m. Calculate the difference of potential between the center and the circumference of the wheel. The declination is 30° west.

3. The potential difference across a generator when delivering 10 amperes is 550 volts. When delivering 40 amps. the potential difference across the generator falls to 547 volts. Find the e.m.f. of the generator and the armature resistance.

4. A 5 Kw. generator when running at 1500 r.p.m. maintains 500 volts across its terminals. The internal resistance of the generator is 0.1 ohm. Find (*a*) the full-load current; (*b*) the e.m.f.; (*c*) the power transformed into heat; (*d*) the efficiency; (*e*) the number of lines of induction cut per second by the armature; (*f*) the number of lines of induction cut per revolution.

5. A generator is connected through 10 ohms resistance to a motor whose internal resistance is 0.6 ohms, the resistance of the generator is 0.4 ohms and its e.m.f. is 110 volts. An ammeter shows only 3 amperes flowing in the circuit. Explain and make all possible computations from the data given.

6. A series-wound motor takes a current of 75 amps. at full load, the applied potential difference being 110 volts and the back e.m.f. 100 volts. The resistance of the armature is 0.04 ohm and resistance of the field coils is 0.02 ohm. Find (*a*) the efficiency, (*b*) the losses other than copper losses in the armature and field coils.

7. (*a*) Find the current in each branch of the circuit shown in Fig. 148. How much power is consumed in the lamp branch? How much in the motor branch?

(*b*) If the load on the motor is lightened, allowing the motor to speed up until the back e.m.f. is 60 volts, what does the current through the lamp become? Explain why this should be so.

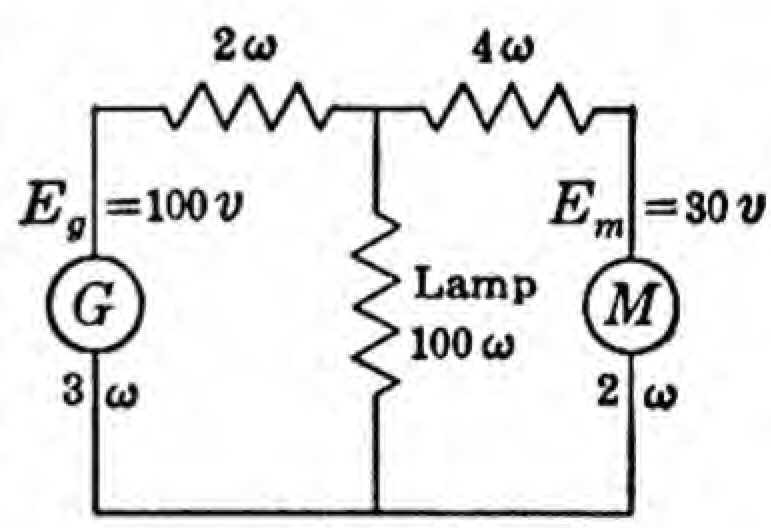


FIG. 148

168. Self Induction. — Associated with every closed electric circuit carrying current there is a definite number of lines of induction. The number of these lines, Φ, is always directly proportional to the current flowing. For a coil of N turns we have

$$N\Phi = li. \tag{253}$$

The proportionality constant l is called the coefficient of self induction or simply the self inductance of the circuit. It is dependent on only the geometrical shape of the circuit. From the above

equation, we define the self inductance of a circuit to be one e.m.u. when one e.m.u. of current flowing in the circuit produces one line of flux linkage.

If the current in a coil is changed, the flux linkage changes, and by Lenz's law an induced e.m.f. is set up which opposes the change. The magnitude of this e.m.f. is

$$e = N\frac{d\Phi}{dt} = l\frac{di}{dt},$$

or

$$E = \frac{N}{10^8}\frac{d\Phi}{dt} = L\frac{dI}{dt}. \tag{254}$$

This gives a second definition of self inductance, namely, the ratio of the induced e.m.f. to the rate of change of the current. The self inductance is one e.m.u. if an induced e.m.f. of one e.m.u. is produced in the circuit when the current changes at the rate of one e.m.u. per sec. The self inductance of a circuit is one henry if an e.m.f. of one volt is induced when the current changes at the rate of one ampere per second. Then from Eq. 253 we see that *a circuit has one henry of self inductance when there are 10^8 lines linked with the circuit for each ampere flowing.* Eq. 253 may be written

$$\frac{N\Phi}{10^8} = LI.$$

In a circuit of one henry of inductance, 10 amps. or one e.m.u. of current would set up 10^9 lines of flux linkage, hence 10^9 e.m.u. of inductance are equivalent to one henry.

The value of l for any given coil may be determined experimentally by means of the induced e.m.f. with the aid of Eq. 254, but when a coil is made carefully to certain exact dimensions the value of l may be calculated by aid of Eq. 253. Thus, consider a long circular solenoid of N turns, radius a, and length b. The magnetic field is constant over the cross section πa^2. Hence

$$\Phi = \mu\frac{4\pi Ni}{b}\cdot\pi a^2,$$

and

$$N\Phi = \mu\frac{4\pi^2N^2a^2}{b}i.$$

Comparing this expression with Eq. 253, one sees that

$$l = \mu\frac{4\pi^2N^2a^2}{b}\text{ e.m.u.},$$

or

$$L = \mu\frac{4\pi^2N^2a^2}{10^9 b}\text{ henries.} \tag{255}$$

The value of l for a circular current is not so easily calculated since the value of the field strength varies over the surface. Formulae for the values of l for all the customary shapes of coils have been derived and may be found in various handbooks.

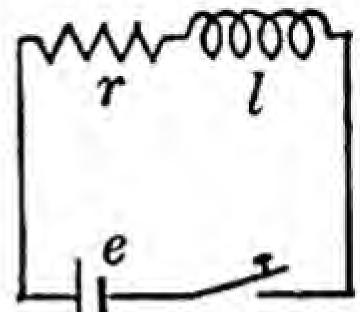


FIG. 149

169. The Growth of Current in a Circuit Containing Resistance and Inductance. — When a series circuit (Fig. 149), containing resistance r, self inductance l, and an e.m.f. e, is closed, the opposing e.m.f. of the inductance allows the current to rise only gradually to its final value $i = e/r$. While the current is increasing at the rate of di/dt, the opposing e.m.f. is $l\,di/dt$. At any instant, Ohm's law gives

$$e - l\frac{di}{dt} = ir. \tag{256}$$

Separating variables,
$$\frac{di}{e - ir} = \frac{dt}{l}.$$

Integrating,
$$-\frac{1}{r}\log(e - ir) = \frac{t}{l} + c.$$

In order to evaluate c, we substitute the initial conditions: when $t = 0$, $i = 0$. Then $c = -1/r \log e$, and

$$\log\frac{e}{e - ir} = \frac{r}{l}t,$$

or
$$\frac{e}{e - ir} = \epsilon^{\frac{r}{l}t},$$

where ϵ is the base of the natural logarithms. Whence

$$i = \frac{e}{r}\left(1 - \epsilon^{-\frac{r}{l}t}\right) = i_m\left(1 - \epsilon^{-\frac{r}{l}t}\right), \tag{257}$$

where $i_m = e/r$ is the maximum value reached by the current.

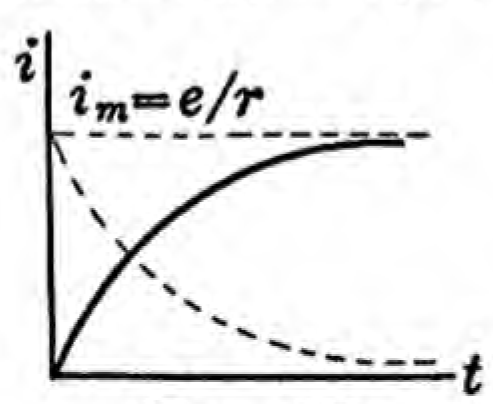


FIG. 150

The relation between i and t in Eq. 257 is shown by the heavy line in Fig. 150. The rapidity with which the current approaches the final maximum value of e/r is determined by the ratio of r/l. If the source of e.m.f. is removed from the circuit and the circuit instantly closed, the current dies out gradually as shown by the dotted curve.

170. The Energy Required to Establish an Electric Current. — As the current i grows, it is continuously doing work against the opposing e.m.f., $l\,di/dt$. The source must supply the energy to

do that work in addition to the energy i^2rt which appears as heat in the resistance. The amount of work done against the induced electromotive force is obtained as follows,

$$dW = ei\,dt = \left(l\frac{di}{dt}\right)i\,dt = li\,di.$$

Integrating from $i = 0$ to $i = i$,

$$W = \int dW = \int_0^i li\,di = \tfrac{1}{2}\,li^2 \text{ ergs},$$

or $$\tfrac{1}{2}\,LI^2 \text{ joules.} \qquad (258)$$

The following analogy may be useful in helping the student understand the effects of self inductance. The self inductance of a circuit corresponds to mass in ordinary dynamics. Mass has the property of inertia which causes an opposition to a change in velocity. In an electrical circuit the self inductance sets up an e.m.f. in opposition to any change in the current in the circuit. When a force has acted on a body and given it a velocity, it has imparted to it energy of an amount $\frac{1}{2}\,mv^2$. An electric circuit which has set up in it a current (and around it a magnetic field) has energy given to it of amount $\frac{1}{2}\,li^2$, as shown above. This energy is associated with the magnetic field and exists in the space surrounding the circuit. When the current is once established, the magnetic field seems to endow it with inertia for, as pointed out above, when the e.m.f. is removed, the current does not die out instantly. As the current starts to "decay," the induced e.m.f., $l\,di/dt$, opposes the decrease and the current continues to flow, although it continually decreases in size until all the energy $\frac{1}{2}\,li^2$, which was stored in the field, is returned and dissipated as heat in the resistance of the circuit.

171. Mutual Induction. — If a current i_p flows in a coil, a portion of the flux associated with that current will pass through or "thread" through any neighboring coil. If the current in the primary is increased, the field strength everywhere is increased in proportion so that the number of flux linkages with a second coil is increased in the same proportion. Hence, using appropriate subscripts to distinguish the primary from the secondary,

$$N_s\Phi_s = mi_p,$$

or $$\frac{N_s\Phi_s}{10^8} = MI_p. \qquad (259)$$

The proportionality constant, m, is called the mutual inductance of the two coils. *We may define the mutual inductance of a pair of circuits to be equal to the total number of linkages of the secondary coil when a unit current flows in the primary coil.*

When the current in the primary circuit changes, there is a change of flux in the secondary and an e.m.f. is produced whose value is

$$e_s = \frac{N_s\, d\Phi_s}{dt} = m\,\frac{di_p}{dt},$$

or

$$E_s = \frac{N_s\, d\Phi_s}{10^8\, dt} = M\,\frac{dI_p}{dt}. \tag{260}$$

Eqs. 260 give a second set of definitions of m and M. Two magnetically linked circuits possess 1 e.m.u. of mutual inductance if an e.m.f. of 1 e.m.u. is induced in the secondary coil due to the current changing in the primary at the rate of 1 e.m.u. per second. The student should be able to make the proper definition of the henry of mutual inductance.

The value of M is easily computed for the case where the primary coil is a long straight solenoid or a closed ring (toroid) of cross sectional area A, with the secondary wound directly on the primary coil. In this case all the flux, Φ_p, of the primary circuit threads through every turn of the secondary coil. From Eq. 217,

$$\Phi_s = \Phi_p = \mu A H = \mu A\,\frac{4\,\pi N_p i_p}{b},$$

where b is the length of the solenoid. If there are N_s turns of the secondary winding,

$$N_s\Phi_s = \mu A\,\frac{4\,\pi N_p N_s}{b}\,i_p.$$

From Eq. 259

$$m = \mu\,\frac{4\,\pi N_p N_s A}{b}\ \text{e.m.u., or}$$

$$M = \mu\,\frac{4\,\pi N_p N_s A}{10^9\, b}\ \text{henries.} \tag{261}$$

It should be noted that, since M depends on the product N_pN_s, the value of M is the same no matter which coil is used as the primary. In fact, the interchangeability of the primary and secondary is a property of any pair of coils.

PROBLEMS

1. Calculate the magnetic flux and flux linkage due to a current in a coil having 1000 turns and an inductance of 0.05 henry when a current of 25 amperes is flowing.

2. Find the inductance of a cylindrical coil, 50 cm. long, 6 cm. diameter, and containing 500 turns. Express the result in millihenries.

3. The coefficient of mutual inductance of an induction coil is 3 henries. If a current of 10 amperes flowing in the primary is reduced to zero in 1/1000 second, what is the e.m.f. at the terminals of the secondary?

4. Making use of § 169, explain why a much higher potential is produced in the secondary of an induction coil upon opening the primary circuit than upon closing it.

5. What is the energy in joules stored in the magnetic field in the case of a circuit having an inductance of 0.02 henries, when a current of 100 amperes is flowing?

6. A coil of 400 turns has a self inductance of 25 millihenries. A second coil of 1000 turns is placed near it so that 20 per cent of the flux of the first coil passes through it. What is the mutual inductance of the two coils?

7. By use of the method of § 163 and Eq. 260, prove that when the current in the primary circuit is increased from zero to i_p, the quantity q_s which is induced in the closed secondary circuit of resistance r_s is given by the expression

$$q_s = \frac{m}{r_s} i_p,$$

or

$$Q_s = \frac{M}{R_s} I_p. \tag{262}$$

8. If two identical flat coils are placed against each other, show that the mutual inductance is numerically the same as the self inductance of either coil taken by itself.

172. Units in the Electrostatic System and Their Dimensions. —

The sole electrical basis for units in the electrostatic system is Coulomb's law of the forces between two static charges. The other physical quantities which enter are purely mechanical quantities of mass, length, and time, and the derived units of force, work, area, etc. We shall proceed to examine the definitions of the electrical quantities and simultaneously derive their dimensions.

Charge — Quantity of Electricity (see § 114). — Coulomb's law of attraction between stationary charges enables us to define the e.s.u. of charge. The dimensional equation for Coulomb's law becomes

$$[MLT^{-2}] = \left[\frac{q^2}{kL^2}\right].$$

Because of our lack of knowledge of why charges attract or repel we cannot express the mechanical dimensions of k. If k is assumed to have no dimensions and is left out of the equation, we obtain:

$$[q] = [M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}].$$

No meaning can be attached to fractional powers of M and L. Besides, we rather expect that since the dielectric medium so modifies the electric forces, k should have dimensions. We do not know what dimensions to give it so we simply leave it in the equations as a required new dimension. Thus we obtain:

$$[q] = [M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}k^{\frac{1}{2}}].$$

It is expected that if it ever becomes possible to express k in mechanical units that the terms in M and L would be rationalized.

In § 119 we defined potential difference as the *work* done on a unit of quantity of electricity. The additional physical quantity involved is the purely mechanical quantity of work. Similarly from Eq. 115 we have defined electric field strength in terms of *force* per unit charge. A current of electricity is obviously the rate of flow of electricity and so is a ratio of the quantity of electricity to the *time* of transfer. The dimensions of these three electrical quantities are obtained as follows:

Potential Difference. —

$$[V] = \frac{[\text{work}]}{[q]} = \frac{[ML^2T^{-2}]}{[M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}k^{\frac{1}{2}}]} = [M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}k^{-\frac{1}{2}}].$$

Electric Field Intensity. —

$$[E] = \frac{[\text{force}]}{[q]} = \frac{[MLT^{-2}]}{[M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}k^{\frac{1}{2}}]} = [M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1}k^{-\frac{1}{2}}].$$

Electric Current. —

$$[i] = \frac{[q]}{[\text{time}]} = [M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2}k^{\frac{1}{2}}]. \tag{263}$$

We may now obtain the dimensions of resistance and capacity —

$$[\text{resistance}] = \frac{[\text{potential difference}]}{[\text{current}]} = \frac{[M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}k^{-\frac{1}{2}}]}{[M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2}k^{\frac{1}{2}}]} = [L^{-1}Tk^{-1}].$$

$$[\text{capacity}] = \frac{[\text{charge}]}{[\text{potential difference}]} = \frac{[M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}k^{\frac{1}{2}}]}{[M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}k^{-\frac{1}{2}}]} = [Lk].$$

Since the only mechanical dimension of capacity is length, the e.s.u. of capacity is often called the centimeter.

We may even obtain the dimensions, in the electrostatic system, of magnetic quantities. From Laplace's equation (Eq. 212) we have

$$dH = \frac{i\,dl \sin\theta}{\rho^2}.$$

Hence
$$[H] = [M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-2}k^{\frac{1}{2}}].$$

The student is expected to work out the dimensions of pole strength, magnetic moment, self inductance, and flux. The results are given below.

$$[m] = [M^{\frac{1}{2}}L^{\frac{1}{2}}k^{-\frac{1}{2}}].$$

$$[\text{magnetic moment}] = [M^{\frac{1}{2}}L^{\frac{3}{2}}k^{-\frac{1}{2}}].$$

$$[l] = [L^{-1}T^{2}k^{-1}].$$

$$[\Phi] = [M^{\frac{1}{2}}L^{\frac{1}{2}}k^{-\frac{1}{2}}].$$

It may be well to repeat that the force between two unit charges is the sole electrical basis of the size of all the above units in the electrostatic system. From that one electrical unit all the others were built up by introduction of additional mechanical units of M, L, and T.

173. Units in the Electromagnetic System and Their Dimensions. — The electromagnetic system of units has its origin in the magnitude of the force between two unit poles, the size selected for the unit poles determining the sizes of all other electric and magnetic units.

Just as was found in the preceding section, we find here that mechanical units alone are not sufficient for expressing the dimensions of magnetic quantities. The permeability of the medium must be taken into account and must appear in all the dimensional equations.

From Coulomb's law we obtain

$$[m] = [M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}\mu^{\frac{1}{2}}].$$

Magnetic Field Intensity. — By introduction of the mechanical quantity of force we obtain:

$$[H] = \frac{[\text{force}]}{[m]} = [M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1}\mu^{-\frac{1}{2}}].$$

Magnetic Moment. —

$$[\text{magnetic moment}] = [M^{\frac{1}{2}}L^{\frac{5}{2}}T^{-1}\mu^{\frac{1}{2}}].$$

Electric Current. — Using Laplace's equation, we obtain the dimensions of current in terms of its magnetic field,

$$i = \frac{dH\rho^2}{dl \sin\theta}.$$

$$[i] = [M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}\mu^{-\frac{1}{2}}]. \tag{264}$$

Quantity of Electricity. —

$$[q] = [i \cdot t] = [M^{\frac{1}{2}}L^{\frac{1}{2}}\mu^{-\frac{1}{2}}].$$

Potential Difference. —

$$[V] = \frac{[\text{work}]}{[\text{charge}]} = [M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2}\mu^{\frac{1}{2}}].$$

Capacity. —

$$[c] = \frac{[q]}{[V]} = [L^{-1}T^{2}\mu^{-1}].$$

Inductance. —

$$[l] = [V]\left[\frac{dt}{di}\right] = [L\mu].$$

Because L is the only mechanical unit which enters into the above expression, the electromagnetic unit of inductance is called the centimeter.

Resistance. —

$$[r] = \frac{[V]}{[i]} = [LT^{-1}\mu].$$

We thus see that all of the electromagnetic units have their foundation on the forces between magnetic poles. From that one magnetic unit all other units were built up with the introduction solely of the mechanical units of M, L, and T.

174. The Relations between the Electrostatic and the Electromagnetic Systems of Units. — For many years there was no known connection between electrostatic and electromagnetic phenomena. After steady currents of electricity were produced by voltaic cells, it was not long before the magnetic fields surrounding them were discovered. Even then, the connection between current electricity and static electricity was not known. It was not until 1889 that Rowland proved that static charges in motion produce a magnetic

field around them. Thus Rowland proved that the difference between static and current electricity is only a question of motion.

Let us now see what information may be obtained from comparing the dimensional formulae of the two systems. Since an electric current is a flow of electricity it is reasonable to assume that its dimensions must be the same regardless of the manner of defining the unit current. Therefore we may equate the dimensional formulae 263 and 264. Thus

$$[M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2}k^{\frac{1}{2}}] = [M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}\mu^{-\frac{1}{2}}].$$

$$\left[\frac{1}{k\mu}\right]^{\frac{1}{2}} = [LT^{-1}].$$

If we equate the dimensional formulae of each of the other electrical quantities, we obtain the same result. We find that $(k\mu)^{-\frac{1}{2}}$ has the dimensions of a velocity.

Although much is known experimentally concerning the general connection between electrical and magnetic phenomena, still enough theory has not been developed that the sizes of the electromagnetic units may be computed from the sizes of the corresponding electrostatic unit. A unit current in the e.s. system is defined as such that one e.s.u. of quantity of electricity is transferred each second. In the e.m. system a unit current is such that when it flows at right angles to a unit magnetic field there is a side thrust of one dyne on each centimeter length of the current. The question of the relative sizes of these units arose early. Experimental measurement showed at once that the e.m.u. was about 3×10^{10} times larger than the e.s.u. The number was significantly close to the magnitude of the velocity of light (§ 187). As the measurement of the ratio was made more and more accurately, the more and more nearly was it found equal to the most accurate determination of the speed of light. Maxwell concluded that since the reciprocal of $\sqrt{k\mu}$ has the dimension of a velocity and since the ratio of the magnitudes of any electrical quantity in the two systems of units is equal to the velocity of light or its square, then certainly light must be an electromagnetic wave. From this idea he developed the foundations of the theory of electromagnetic radiation.

a ✓

175. The Practical Units. Legal Units. General Relations among the Three Systems of Units. — It was found that most of the e.m. or e.s. units were either too large or too small to be used

conveniently, so certain decimal fractions or multiples of these units were chosen as the so-called Practical System. The e.m.f. of a voltaic cell was thought to be of a practical size for a unit. It was slightly larger than 10^8 e.m.u.'s of potential difference. Therefore the *volt was defined to be equal exactly to 10^8 e.m.u. of potential difference.* On this basis a voltaic cell (zinc and copper in dilute sulphuric acid) has an e.m.f. of about 1.10 volts. The current which was thought to be practical in size was called the ampere. *The ampere was defined as exactly one tenth of an e.m.u. of current.* The quantity of electricity transmitted in one second by a current of one ampere was called the coulomb. Hence, when one coulomb of electricity is transferred across one volt of potential difference, $10^8 \times \frac{1}{10}$ or 10^7 ergs of work is done. In order to do away with such a large number, *10^7 ergs is called one joule.* It is seen that the practical system is still a decimal c.g.s. system and differs from the e.m. system only by powers of 10.

Since by experiment 3×10^{10} (nearly) e.s.u. of current is equal to 1 e.m.u. of current, we have the first line of the following table.

	E.S.U.	E.M.U.	PRACTICAL
Q or I	3×10^{10}	1	10
V	1	3×10^{10}	300 volts

Thus

10 amperes = 1 e.m.u. of current = 3×10^{10} e.s.u. of current.
10 coulombs = 1 e.m.u. of quantity = 3×10^{10} e.s.u. of quantity.

The student is expected to show that from the definitions of potential difference in the e.m. and e.s. systems, it follows that 1 e.s.u. = 3×10^{10} e.m.u. of potential difference. From the above definition of the volt, 1 e.s.u. = 300 volts.

With the aid of this easily memorized table the relations in the three systems of all other electrical quantities may be found. Thus,

$$1 \text{ farad} = \frac{1 \text{ coulomb}}{1 \text{ volt}} = \frac{\frac{1}{10} \text{ e.m.u. of } q}{10^8 \text{ e.m.u. of } v} = \frac{1}{10^9} \text{ e.m.u. of capacity}$$

$$= \frac{3 \times 10^9 \text{ e.s.u. of } q}{\frac{1}{300} \text{ e.s.u. of } v} = 9 \times 10^{11} \text{ e.s.u. of capacity.}$$

It is seen that the e.s.u. of capacity (the centimeter) is very small and the e.m.u. is very large. The farad of capacity although expressed in the practical units of coulombs and volts is entirely too large for practical use, so one millionth of it, or the microfarad, is the customary unit. The student is expected to work out the equivalents of the henry and the ohm in the other systems of units.

Before the various countries of the world had established their governmental bureaus for measuring and intercomparing electrical standards, and before commercial companies had been organized to manufacture electrical apparatus of guaranteed accuracy, a need was felt for specifications as to how to set up these standards. Various specifications were made and legalized only to be discovered in error as experimental technique improved. Finally there were international congresses appointed to set up new and more accurate specifications. In turn these were made obsolete by the ever advancing skill of experimenters. Thus the International Ohm of 1893 was defined as the resistance to a steady current, of a column of mercury of uniform cross section, having a length of 106.300 cm. and a mass of 14.4521 gm., when at a temperature of 0° C. In 1908 additional specifications were made concerning the cross section of the mercury column, and particularly concerning the manner of getting electrical connection onto the ends of the column. The International Ampere was defined in terms of the amount of silver deposited electrolytically in a given time. The International Volt was defined by Ohm's law in terms of the International Ampere and Ohm. A committee was appointed to measure the e.m.f. of the Weston Standard cell (made according to elaborate specifications). The experimental value of such a cell in terms of the International Ampere and Ohm, which the committee reported, has now become a second definition of the International Volt. As soon as greater precision was obtainable in electrical measurements, this experimentally defined volt was destined to become inconsistent with the International Volt.

Thus the number of legalized units multiplied until the matter became objectionable. So, in the International Congress of Weights and Measures, in 1933, all legal units were recommended to be dropped. From now on, there are to be considered only the metric units of the electromagnetic system. The practical units are only certain fractions or multiples thereof. This procedure

has only recently been made possible by the co-operation of bureaus of standards of England, France, Germany, U. S. S. R., Japan, and the United States in maintaining regular intercomparison of standards. An accuracy of 1 part in 100,000 is expected to be obtained shortly for the three fundamental units: the ohm, volt, and ampere.

The Ballistic Galvanometer

176. The Equation for a d'Arsonval Galvanometer Used Ballistically. — In § 157 we studied the effect of a steady current passing through a d'Arsonval galvanometer. In this paragraph we shall study the effect of a momentary current. Let a condenser be charged by attaching a battery of a certain voltage across its terminals. When the terminals of the condenser (disconnected from the battery) are connected across the terminals of the galvanometer, the displaced electrons move to their equilibrium positions, producing a temporary current through the galvanometer coil. At the instant when the charged condenser is connected to the galvanometer, the current is zero but the current rapidly increases to a maximum and then begins to decrease as the condenser is discharged, finally reaching zero. The whole time of discharge may have occurred in a thousandth part of a second or less. While the current is flowing, a torque acts upon the coil as calculated in § 156. This torque, though lasting only a short time, imparts to the coil a certain angular momentum $J\omega = \int L\,dt$, according to Eq. 75. (It is necessary in this section to use J for moment of inertia to avoid confusion with I which is used for the current.) The angular velocity of the coil is increased from 0 to ω in such an exceedingly short time that the angle turned through during that time, $\theta = \frac{1}{2}\alpha t^2$, is usually too small to be appreciable when compared to its subsequent motion. For all practical purposes we may think of the coil as starting from its rest position with an angular velocity ω as if each side of the coil were suddenly projected from a gun. Hence it is seen why a galvanometer so used, for electrical impulses of short duration, should be called a "ballistic" galvanometer. The coil continues to swing from side to side through decreasing angles until friction finally brings it to rest.

Suppose that the secondary coil of a mutual inductance is connected across the terminals of a galvanometer (Fig. 152). When the primary circuit is closed, the momentary flow of current in the secondary which passes through the galvanometer resembles the discharge of the condenser, rising to a maximum and dying out to zero in a very short time. The quantity of electricity which has flowed through the coil is given by Eq. 262.

We shall now show that, under the conditions described, the maximum angle of the first swing of the galvanometer is always proportional to the quantity of electricity passing through the galvanometer. We shall derive the equation, assuming that there are no frictional forces to damp out the vibrations, and we shall later employ means for correcting for the damping which always occurs. By Eqs. 75 and 241,

$$J(\omega - 0) = \int L\,dt = \int NiAH\,dt = NAHq. \qquad (265)$$

This equation tells us that the angular velocity with which the coil begins its swing is directly proportional to the quantity of electricity which has passed through it. Since ω cannot easily be measured, we eliminate it by using an energy equation. The coil starts its swing with a certain amount of kinetic energy, $\frac{1}{2}J\omega^2$. At the end of the swing, ω is zero and all the previous kinetic energy is stored up as potential energy in the twisted suspension. The potential energy may be found by adding up (integrating) the work done in twisting the suspension through each infinitesimal angle. Let L_0 be the torque necessary to twist the suspension through one radian. For a twist of θ radians, the torque $L_0\theta$ is required. The work required to twist the suspension from θ to $\theta + d\theta$ is $L_0\theta\,d\theta$, and the total work from zero twist is given by

$$W = \int_0^\theta L_0\theta\,d\theta = \tfrac{1}{2}L_0\theta^2 = \tfrac{1}{2}J\omega^2.$$

Solving for ω and substituting in Eq. 265, we have

$$q = \frac{J}{NAH}\omega = \frac{J}{NAH}\sqrt{\frac{L_0}{J}}\,\theta = \frac{L_0}{NAH}\sqrt{\frac{J}{L_0}}\,\theta.$$

This latter algebraic step is made because we wish to have L_0 instead of J appear in the final formula.

As given in Eq. 92 the period of free vibration of the galvanometer coil is

$$T = 2\pi\sqrt{\frac{J}{L_0}}.$$

Hence,
$$\left.\begin{aligned} q &= \frac{L_0 T}{2\pi NAH}\,\theta = q_0\theta, \\ Q &= Q_0\theta. \end{aligned}\right\} \qquad (266)$$
or

It is seen that q_0 is the number of e.m.u. of quantity required to produce one radian displacement of the galvanometer coil and Q_0 the number of coulombs required for the same displacement.

Thus for an impulsive current which becomes zero before the galvanometer coil has time to move appreciably, and for the case of no losses through friction, the maximum deflection of the coil is directly proportional to the quantity of electricity. The quantity q_0 is called the ballistic constant of the instrument.

In Eq. 242, it will be noticed that the current constant, k, of a d'Arsonval galvanometer is L_0/NAH. So we have the relation

$$k = \frac{2\pi}{T}\,q_0.$$

The constant k is easily determinable, as is also T, and then q_0 may be computed. This calculated value of q_0, however, is for completely undamped motion and so its value has to be further modified before it can be used practically.

177. The Effect of Damping in Simple Harmonic Motion. — When there is no friction present, the return torque of the suspension is the only action on the coil. The torque is always directed oppositely to the displacement and is directly proportional to it; hence the coil undergoes angular harmonic motion, the equation of motion being

$$K\frac{d^2\theta}{dt^2} = -L_0\theta.$$

Besides the action of the suspension on the coil, there is air friction, a reaction due to eddy currents in the frame and windings of the coil or in the "damping coil" hung on the frame, and also a reaction due to the induced current which flows through the coil and the external circuit (in case there is a resistance shunt on the galvanometer or in case the galvanometer is connected to the secondary of a mutual inductance). For small velocities, the fric-

tion of the air on a moving body is directly proportional to the first power of the velocity. When the coil is in a radial field, any induced e.m.f. will be directly proportional to the angular velocity of the coil. The induced currents will be directly proportional to the e.m.f. and the opposing torque will be directly proportional to the current. Lenz's law states that the induced current will flow in such a direction as to oppose the motion. Hence the torque produced by the induced currents will be proportional to the angular velocity and oppose the motion, — just as does the air friction. Let A be the torque due to air friction and induced currents when the angular velocity is one radian per second. Then we may write Newton's second law as follows,

$$- L_0\theta - A\frac{d\theta}{dt} = K\frac{d^2\theta}{dt^2}.$$

The solution of this second order differential equation with constant coefficients is

$$\theta = \theta_0\epsilon^{-at}\sin\left(\frac{2\pi}{T}t\right), \tag{267}$$

where ϵ is the base of the Naperian logarithms, T the period of the motion, and θ_0 and a are constants. The solution shows that the motion is still periodic but the amplitude decreases exponentially with the time. The graph of Eq. 267 is shown in Fig. 151. The first deflection of the galvanometer is represented by θ_1. Following that, there is a smaller deflection θ_2 in the opposite direction, θ_3 in the original direction, etc. If the damping is small, the curve passing the peaks of the deflection-time curve is given approximately by the equation

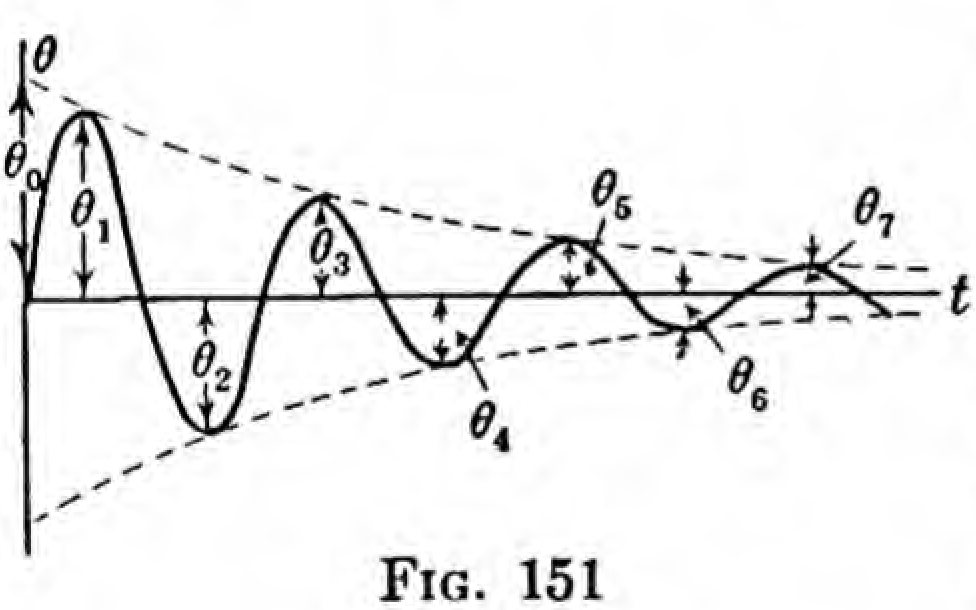


FIG. 151

$$\theta = \theta_0\epsilon^{-at}.$$

The angle θ_0 would have been reached on the first swing if it had not been for the damping of the motion. That value is represented by the ordinate where the dotted curve crosses the θ axis.

We shall now show how θ_0 may be computed, having given θ_1, θ_2, θ_3, etc. From Eq. 267, we obtain the following:

$$\left.\begin{array}{lll}\text{when} & t = 0, & \theta = 0. \\ & t = \dfrac{T}{4}, & \theta_1 = \theta_0 \epsilon^{-\frac{aT}{4}}. \\ & t = \dfrac{3\,T}{4}, & \theta_2 = \theta_0 \epsilon^{-\frac{3\,aT}{4}}. \\ & t = \dfrac{5\,T}{4}, & \theta_3 = \theta_0 \epsilon^{-\frac{5\,aT}{4}}. \\ \cdot & \cdot\quad\cdot\quad\cdot\quad\cdot & \cdot\quad\cdot\quad\cdot\quad\cdot \\ & t = \dfrac{2\,n-1}{4}\,T, & \theta_n = \theta_0 \epsilon^{-\frac{(2\,n-1)aT}{4}} \\ & t = \dfrac{2\,n+1}{4}\,T, & \theta_n + 1 = \theta_0 \epsilon^{-\frac{(2\,n+1)aT}{4}} \end{array}\right\} \tag{268}$$

Hence, $$\frac{\theta_n}{\theta_{n+1}} = \epsilon^{\frac{aT}{2}} = \frac{\theta_1}{\theta_2} = \frac{\theta_2}{\theta_3} = \frac{\theta_3}{\theta_4} = \cdots = \rho. \tag{269}$$

This ratio ρ is called the *damping ratio.* From the second line in Eqs. 268, we obtain

$$\theta_0 = \theta\epsilon^{\frac{aT}{4}} = \theta_1\sqrt{\rho}. \tag{270}$$

Hence, if we know the damping ratio ρ, we may correct the first deflection θ_1 for the effect of damping.

By using the ratio of two successive values of θ, it is not usually possible to get an accurate determination of ρ. An accurate value may be gotten as follows: From Eq. 269

$$\frac{\theta_1}{\theta_2} \cdot \frac{\theta_2}{\theta_3} = \rho^2 = \frac{\theta_1}{\theta_3}.$$

Likewise, $$\frac{\theta_n}{\theta_{n+m}} = \rho^m,$$

or $$\rho = \sqrt[m]{\frac{\theta_n}{\theta_{n+m}}}. \tag{271}$$

When the damping effect is small, an approximate correction may be made. Let λ represent the logarithm to the base ϵ of the ratio of two successive deflections. This quantity is called the *logarithmic decrement.*

Hence, $$\lambda = \log_\epsilon \frac{\theta_1}{\theta_2} = \log_\epsilon \frac{\theta_2}{\theta_3} = \cdots \log_\epsilon \rho = \frac{aT}{2}. \tag{272}$$

From the second line of Eqs. 268,

$$\theta_0 = \theta_1\epsilon^{\frac{aT}{4}} = \theta_1\epsilon^{\frac{\lambda}{2}} = \theta_1\left(1 + \frac{\lambda}{2} + \text{higher terms}\right). \tag{273}$$

When the damping is small, the higher terms are negligible. From Eq. 271, we obtain

$$\log_\epsilon \frac{\theta_n}{\theta_{n+m}} = m \log_\epsilon \rho = m\lambda.$$

So
$$\lambda = \frac{1}{m} \log_\epsilon \frac{\theta_n}{\theta_{n+m}} = \frac{2.303}{m} \log_{10} \frac{\theta_n}{\theta_{n+m}}. \tag{274}$$

It can be shown that the error in the value of ρ due to error in observing θ_n and θ_{n+m} is the least when a sufficient number of oscillations have elapsed to make

$$\theta_n = \epsilon\, \theta_{n+m}.$$

178. The Calibration of a Ballistic Galvanometer. — The simplest means of calibrating a ballistic galvanometer (determination of the constant Q_0) is to send a known quantity of electricity through the galvanometer and observe the deflection produced. Thus if a condenser of capacity C is charged from a standard cell of e.m.f. V and is discharged through the galvanometer, then $Q = CV$ and Eqs. 266 and 270 give

$$Q_0 = \frac{CV}{\theta_1\sqrt{\rho}}, \quad \text{or} \quad \frac{CV}{\theta_1\left(1 + \frac{\lambda}{2}\right)}. \tag{275}$$

As already mentioned toward the end of § 176, we may compute Q_0 from the current constant of the galvanometer. We have $Q_0 = (T/2\,\pi)K$, where K is the current in amperes required to produce unit deflection. This method has the obvious advantage that K is computed from steady deflections which can be measured accurately and T may be measured accurately. In the first method the deflection θ_1 must be read just as the maximum deflection is reached. Catching a reading "on the fly" on a ballistic galvanometer never allows as great precision as reading steady deflections. In the equation, T is the period of the coil without damping. However, only T_λ, the period with air damping, may be observed. Further theory shows that

$$T = T_\lambda \frac{\pi}{\sqrt{\pi^2 + \lambda^2}}.$$

Usually λ is never more than a few hundredths, so T and T_λ are practically the same. Nevertheless, the difference must be considered until it is experimentally known to be negligible.

A third method of determining Q_0 makes use of a mutual inductance of known dimensions. The galvanometer is connected in series with a resistance and the secondary winding of N_s turns of a mutual inductance (Fig. 152). Let the primary, a long solenoid of length L and with N_p turns, be connected in series with an ammeter A, a battery, and a controlling resistance. When the switch S is closed, the current in the primary increases from zero to I_p. During the time in which the primary current is rising to its value of I_p, there is an induced secondary current. The quantity of electricity transferred by the secondary current is given by Eq. 262. Using Eq. 261 also, we obtain

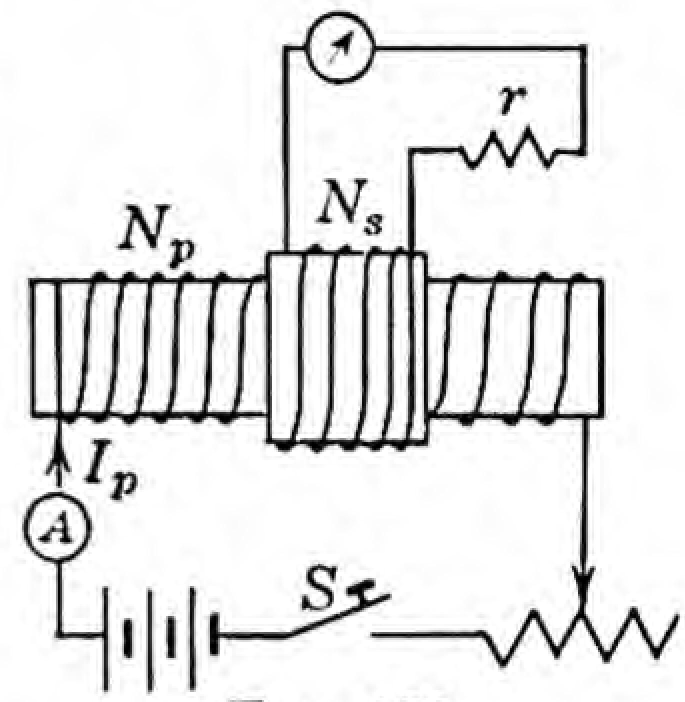


FIG. 152

$$Q_s = Q_0\theta_1\sqrt{\rho} = \frac{M}{R_s} I_p = \frac{4\,\pi N_p N_s A}{10^9\,LR_s} I_p. \tag{276}$$

Then
$$Q_0 = \frac{4\,\pi N_p N_s A}{10^9\,LR_s} \cdot \frac{I_p}{\theta_1\sqrt{\rho}}, \tag{277}$$

where R_s includes the total resistance of the secondary circuit, namely, the resistance of the galvanometer, the windings of the secondary coil, and the controlling resistance r. The galvanometer coil while moving through θ_1 (mostly after Q_s has flowed) cuts lines of force, causing an induced current through R_s. The damping thereby produced is small if R_s is large, and conversely.

This latter method is the least satisfactory of the three unless a mutual inductance is available whose coefficient is very large. If M is small, then even after I_p has been increased as much as possible, θ_1 may be too small. If r is made smaller in order to increase the value of θ_1, then the damping factor ρ becomes large. When ρ becomes large, the vibrations of the galvanometer die out so quickly that ρ cannot be determined accurately. On the other hand, if r is increased, in order to reduce ρ, then the deflection θ_1 is decreased so much that it cannot be measured with accuracy. Some intermediate value of r must be chosen to allow the errors in θ and ρ to be about equal. However, if M is large, then r may be large so that there is not too much damping and there is still produced sufficiently large values of θ_1 that they can be read with accuracy.

Magnetic Induction in Iron

179. Experimental Methods of Measurement of Induction in Ferro-magnetic Substances. — The student should review carefully §§ 128, 131, and 136 before proceeding.

I. The Magnetometer Method. — The magnetometer method of measuring magnetic induction is a simple modification of the method of comparing two magnetic fields as described in § 131. The permanent magnet at O in Fig. 104 is replaced by the sample of magnetic material whose state of magnetization is to be measured. The modified arrangement is pictured in Fig. 153. At P is

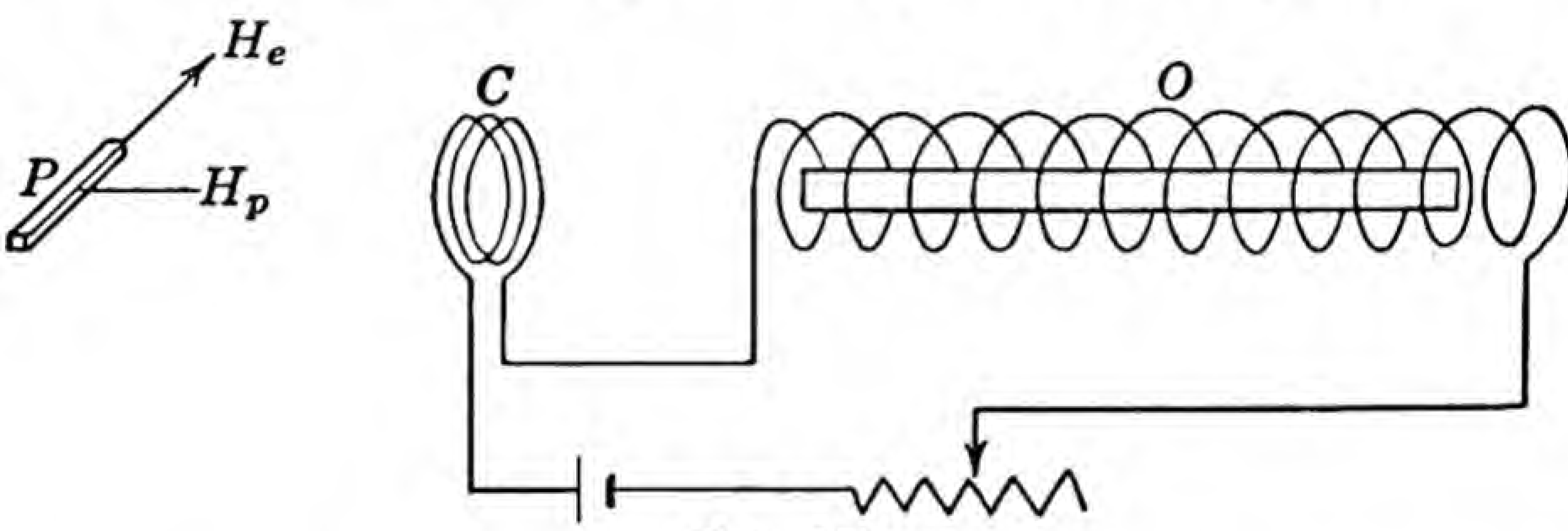


FIG. 153

suspended the small magnetometer needle and some device is employed for measuring its angular deflection from the magnetic north. The sample is placed inside of a long solenoid at O, a point magnetically east (or west) of the suspended needle. A current sent through the solenoid will magnetize the sample. The field at P in the east or west direction will be due to the magnetization of the iron and, in a smaller degree, to the field of the solenoid. To avoid this stray field from the solenoid, there is placed in series with the solenoid a coil C of a few turns which is placed near to P so that its field neutralizes the field due to the empty solenoid (*i.e.* with the magnetic sample removed). The coil C is moved to such a point that when the current in the empty solenoid is either made or broken, the magnetometer needle does not move. When the fields due to C and the solenoid neutralize each other for any one value of current used, then of course neutralization will exist for all values of the current. Under these conditions, when the sample is placed in the solenoid, the resultant field H_p at P is just that due to the induced poles of the sample. For any state of magnetization, when there are m unit poles induced in the sample of cross section a, length $2\,l$, and whose center is distant r from P, we have according to Eqs. 195 and 199,

$$\tan\theta = \frac{H_p}{H_e} = \frac{4\,mlr}{H_e(r^2 - l^2)^2}. \tag{278}$$

Recalling that $m/a = \mathcal{J}$, the intensity of magnetization, we obtain

$$\mathcal{J} = \frac{m}{a} = \frac{H_e(r^2 - l^2)^2 \tan\theta}{4\,alr}. \tag{279}$$

The value of the induction B is obtained from Eq. 204. The value of H may be calculated by Eq. 217. Hence,

$$B = \frac{4\,\pi\,NI}{10\,L} + \frac{\pi H_e(r^2 - l^2)^2 \tan\theta}{alr}. \tag{280}$$

In addition to the mechanical measurements of the dimensions of the solenoid and sample and the position and deflection of the magnetic needle, the value of the earth's field must be measured at P.

In order that the demagnetizing field in the sample shall be small and that the poles shall be very near the ends of the sample, the sample is usually made into a long thin rod. Then l^2 is seldom negligible compared to r^2. Since the magnetic field due to the current in a solenoid is smaller near the ends than the amount $4\,\pi NI/10\,L$ (which assumes infinite extent of the solenoid on each side of the point considered) the solenoid must be built so as to be considerably longer than the samples. Then all parts of the samples will be in a field of the computed magnitude.

II. The Rowland Ring Method. — The sample to be tested by Rowland's method must be welded or cast into a circular ring of either circular or rectangular cross section. Over the whole length of this ring is wound uniformly one or more layers of wire of N_p turns (Fig. 154). (For convenience in drawing, the primary windings are shown in the figure over only a portion of the ring.) A secondary circuit of N_s turns is wound directly over any portion of the primary. This secondary coil is connected in series with a galvanometer, a resistance r, and the secondary of a standard mutual inductance. Whenever the current I_p is varied, there is a change in the flux in the sample which causes a certain quantity of electricity to flow through the galvanometer.

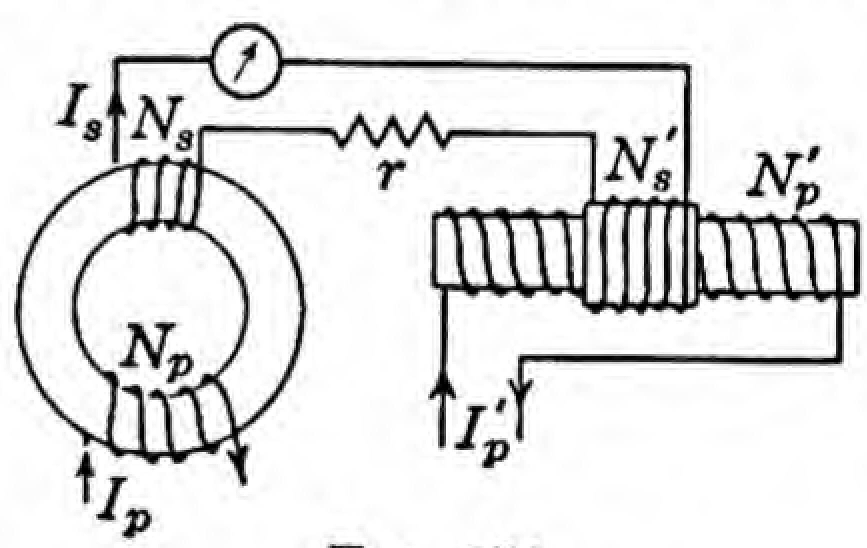


FIG. 154

$$Q = Q_0\theta_1\sqrt{\rho} = \int I_s\,dt = \int \frac{E_s}{R_s}\,dt, \tag{281}$$

where I_p is the variable current induced in the secondary, E_s is the corresponding induced e.m.f. and R_s is the resistance of the galvanometer, both secondary coils, and r, added together. By Eq. 260, we obtain

$$Q_0\theta_1\sqrt{\rho} = \int \frac{E_s}{R_s}\,dt = \int \frac{N_s\,d\Phi}{R_s\,10^8\,dt}\,dt$$
$$= \frac{N_s}{10^8\,R_s}\int A\,dB = \frac{N_sA}{10^8\,R_s}\,\Delta B, \tag{282}$$

where B is the induction in the iron, *i.e.* the lines per square centimeter, $A\,dB$ is the infinitesimal change in flux across the whole cross section A of the iron sample, and ΔB is a finite change in B. Then

$$\Delta B = \frac{Q_0R_s\,10^8}{N_sA}\,\theta_1\sqrt{\rho}. \tag{283}$$

The value of Q_0 is easily obtained by an auxiliary observation as follows: Send current I_p' through the primary of the solenoid which is connected as part of the circuit. By the method described in the previous section, we obtain the value of Q_0 as given in Eq. 277.

The object of having both secondary coils permanently connected in series with the galvanometer may now be seen. In Eq. 277 appears the correction for damping, $\sqrt{\rho}$. A damping factor appears also in Eq. 283. If these two damping factors are the same, they will cancel and the determination of them is avoided. The two values of ρ will be the same unless the resistance in the secondary circuit is changed. So by leaving the two secondary coils connected in the circuit throughout the experiment we eliminate the measurement of the damping ratio.

Substituting the value of Q_0 from Eq. 277 into Eq. 283, using primed letters to refer to the standard solenoid, we get

$$\Delta B = \frac{4\,\pi N_p'N_s'A'}{10\,N_sAL'}\,\frac{I_p'}{\theta'}\,\theta_1. \tag{284}$$

All terms in this equation are constants except θ_1. A study of Eq. 277 will show that I_p'/θ_1' must be a constant. Thus any deflection θ_1 which occurs when the current I_p is changed, indicates a change in the induction in the iron, which is directly proportional to it. The value of the magnetizing field which produces the final value of B is

$$H = \frac{4\,\pi NI}{10\,L},$$

where L is the length of the ring as computed from its mean diameter.

The Rowland ring method has the following disadvantages:

(a) It does not indicate the exact state of magnetization but gives only changes in induction.

(b) The sample must be made into a ring.

(c) The ring must have a uniform winding made over its whole length. It is thus not adapted to rapid tests of a large number of samples.

Modifications of this method have been made which allow straight bars of sample material to be used and allow quick interchange of samples. These modified methods are used now almost to the exclusion of other methods.

The magnetometer has the advantage of indicating by its deflection the exact state of magnetization. Its disadvantages are:

(a) The sample must be a long thin rod.

(b) The earth's field must be measured.

(c) Changes in the local earth's field cause the deflections to vary, — such as caused by passing street cars or by the transfer of any iron mass about the premises.

180. The Measurement of Flux with a Ballistic Galvanometer. — In Fig. 154, suppose that instead of the secondary winding on the Rowland ring, we connect into the galvanometer circuit a movable rigid test coil of N_t turns and of cross-section A_t. Let this coil be placed in a magnetic field whose induction B is to be measured. After the galvanometer is brought to rest the coil is suddenly flipped out of the magnetic field to a place where the field is known to be practically zero, N_tBA_t lines of induction being cut. This total motion must be completed before the galvanometer has had time to move appreciably. The resulting throw of the galvanometer θ_t determines the quantity of electricity Q_t induced in the secondary circuit. Thus from Eq. 250,

$$Q_t = Q_0\theta_t\sqrt{\rho} = \frac{N_tBA_t}{10^8\, R_s},$$

where R_s is the total resistance of the secondary circuit, including the test coil. If the mutual inductance is used to get the value of Q_0, then combining Eq. 277 with the above equation, we have (using the notation in Eq. 284),

$$B = \frac{4\,\pi N_p'N_s'A'}{10\,L'N_tA_t}\,\frac{I_p'}{\theta'}\,\theta_t = K\theta_t. \tag{285}$$

Thus after the constant K has been once determined, the measurement of induction or field strength may be quickly made. In order that the lead wires from the test coil may not enclose lines of force which would be cut when the coil is moved, the leads are made of flexible wire and are twisted closely together.

Suppose the field strength is desired in small air gaps around pole pieces of a motor or other electric machine. It becomes impossible to flip the coil quickly out of such narrow spaces. In such cases the coil is left in place and the magnetizing current in the machine is reversed. The flux reverses and the deflection of the galvanometer is twice what it would have been had the coil been flipped out. Then $B = H = K\theta_t/2$. If, in getting the value of Q_0, the current in the primary of the mutual inductance had been reversed in getting θ', then Eq. 285 would be directly applicable since both θ' and θ_t would have been doubled.

In certain portable fluxmeters, the ballistic galvanometer coil is mounted so that practically no return torque exists upon it. The test coil is connected directly across the terminals of the meter. A series resistance inside the meter is adjusted so that the motion of the galvanometer coil is highly damped. As the test coil is removed from the field, the pointer attached to the galvanometer coil moves up to its maximum position and remains there. The scale is made to read directly in lines per square cm. or lines per square inch. The final reading is independent of the speed of motion of the test coil so long as the speed is not very low. The galvanometer coil must be returned to its zero by mechanical means or by inserting the test coil back into the field until the deflection is zero, opening the galvanometer circuit, and removing the coil.

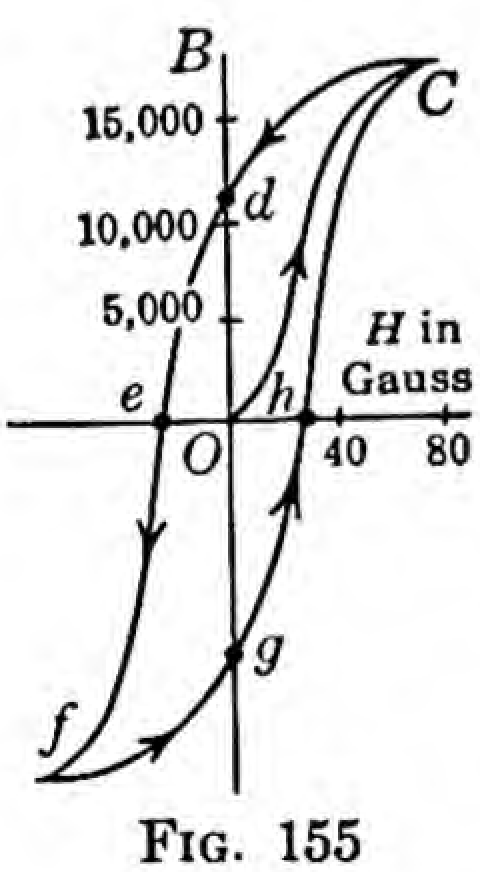


FIG. 155

181. Magnetization Curve and Hysteresis Loop. — It is customary to plot the values of B along the vertical axis and the values of H along the horizontal axis as shown in Fig. 155. When the sample is completely demagnetized, its state is represented by a point at the origin. As the magnetizing field H is applied and increased, the induction varies as shown by the curve OC. As H is increased from zero, the induction B at first increases slowly but later increases very rapidly. Finally the change becomes less

rapid until the increase in B beyond the point C is but little more than the increase in H. At C the sample is said to be *saturated.* The curve OC is called *the magnetization curve.*

If, after the sample is saturated, the field is decreased to zero, the curve Cd is obtained. The values of B do not decrease to the corresponding values reached when the field was being increased. The induction is said to *lag* behind the magnetizing field. The induction represented by Od is called the *residual magnetism* or the *retentivity.* When the current in the magnetizing coil is reversed and increased, at e the induction is reduced to zero. The value of H at the point e is called *the coercive force.* At f the sample is saturated in the reverse direction. Reducing the field to zero, the point g is reached. Reversing the field, saturation is again reached at C. The sample has been carried through a complete cycle of magnetization. After a sample has been carried through several cycles, then the portion $Cdef$ is found to be exactly symmetrical to the portion $fghC$, showing that there is no directional effect. The magnetization takes place in any one direction exactly as it does in the opposite direction. The lag of B behind H is called *hysteresis* and the complete curve $CdefghC$ is called the *hysteresis loop.* It is not uncommon for the curve from h to C to cut across the magnetization curve and approach C from above it.

182. The Work Required per Unit Volume to Magnetize a Specimen. — Let us consider a long straight solenoid of N turns, of cross-sectional area a, and length L which carries a given current i. Let a piece of magnetic material having those same dimensions be inserted into the solenoid. As it is inserted it will become magnetized and the total flux passing through the solenoid will be greatly increased. This change in flux will cause an electromotive force e to be induced in the solenoid which will oppose the flow of current. In order to maintain the current i constant during the insertion of the magnetic material, the e.m.f. in the circuit must be increased by an amount e. Let $d\Phi$ be the change of flux occurring during the time dt. The extra electrical energy, $dW = ei\,dt$ ergs, which is supplied to the solenoid to keep its current unchanged, must be the energy required to produce the increased magnetization. We know that

$$e = N\frac{d\Phi}{dt}, \quad \text{and} \quad H = \frac{4\,\pi N i}{L}.$$

By substituting these values we may change the expression for the work so that it involves only those variables which relate to the magnetized material.

$$dW = ei\,dt = N\frac{d\Phi}{dt}\cdot\frac{HL}{4\,\pi N}\,dt = \frac{HL\,d\Phi}{4\,\pi}.$$

Remembering that $d\Phi = a\,dB$ and $aL = v$ the volume of the material, we obtain

$$dW = \frac{v}{4\,\pi}H\,dB,$$

and

$$W = \frac{v}{4\,\pi}\int_{B_1}^{B_2} H\,dB \text{ ergs.} \tag{286}$$

Now if the induction in the sample has increased from B_1 to B_2 through the values as shown by the curve CC' (Fig. 156), $H\,dB$ indicates the infinitesimal shaded area shown, and

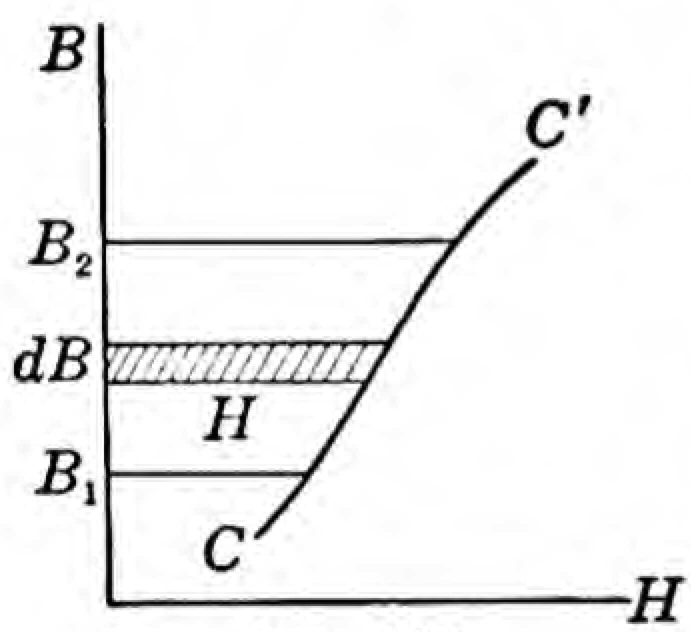


FIG. 156

$$\int_{B_1}^{B_2} H\,dB$$

indicates the area between the curve, the B axis, and the two values of B_1 and B_2. We may then state that the energy required for each cubic centimeter of the sample to change the induction from B_1 to B_2 is the indicated area divided by $4\,\pi$. When either H or dB is negative, the area is negative. If both H and dB are negative, the area is positive. Thus when the magnetization is changed from C' to C, although H decreases it remains positive, while B decreases so that dB is negative. During this change the induced e.m.f. aids the maintenance of the current i and energy is delivered back from the magnetic field of the medium to the electric circuit. Thus, in Fig. 157, as the magnetization is changed from a to c, the energy gained from the field for each cubic centimeter of the material is

$$-\frac{\text{area }(abca)}{4\,\pi};$$

from c to e the work required is

$$+\frac{\text{area }(cdefgc)}{4\,\pi};$$

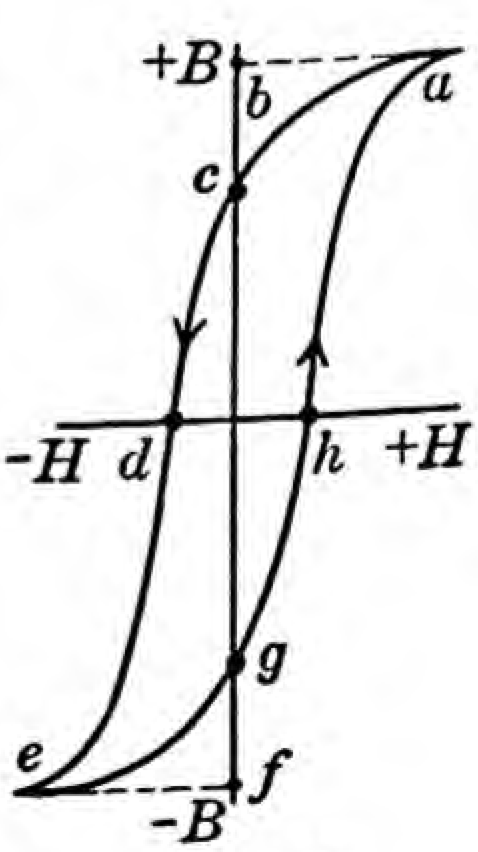


FIG. 157

from e to g the energy gained is

$$-\frac{\text{area }(efge)}{4\pi};$$

from g to a the work required is

$$+\frac{\text{area }(ghabcg)}{4\pi}.$$

Upon adding these expressions, it is seen that the areas $abca$ and $efge$ cancel and only the area enclosed by the loop remains. Hence we may state that the energy required to carry each cubic centimeter of a sample through a complete magnetic cycle is equal to the area of the hysteresis loop divided by 4π. This work is done against the molecular forces and appears as heat in the material. In all alternating current generators and motors where large masses of iron are carried through many cycles of magnetization each second, the heating becomes important and the machines must be designed to radiate the heat due to that and other causes. If a piece of iron of volume v cm³. is magnetized by a current of frequency f, then the heat developed per second due to hysteresis is given by:

$$\text{Heat per second in calories} = \frac{v \times f \times \text{Area of loop}}{4\pi \times 4.2 \times 10^7}.$$

183. The Magnetic Circuit. Magnetomotive Force. — Consider a closed magnetic circuit, such as shown in Fig. 158, made of many kinds of materials of various lengths and cross sections. Let a solenoid of N turns be wound on the ring so that there is no magnetic leakage (*i.e.* no free poles appear and the same flux passes through each section). Consider a unit pole carried once around this circuit (inside the magnetic material). In § 145 we have seen that the work required is

$$W = \frac{4\pi NI}{10}.$$

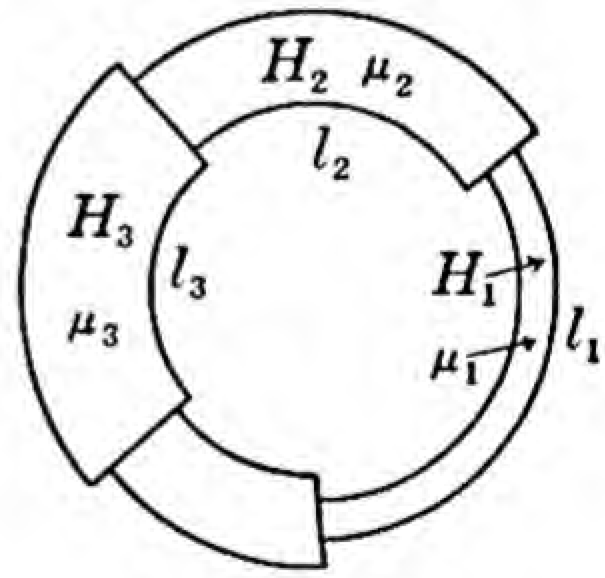


FIG. 158

This work may be computed also in terms of the field strength in the various materials. Thus

$$W = \int H\,dl = H_1l_1 + H_2l_2 + \cdots \tag{287}$$

So

$$NI = \frac{10}{4\pi}(H_1l_1 + H_2l_2 + \cdots). \tag{288}$$

This is the fundamental equation of the magnetic circuit.

From § 149 we see that the amount of work done in transferring a unit quantity of electricity once around a circuit is the electromotive force. By analogy we define the work done in transferring a unit pole once around a magnetic circuit to be the *magnetomotive force*, m.m.f. It is equal to $(4\pi NI/10)$ ergs/pole. The product NI of the number of amperes by the number of turns is called the number of ampere-turns. So $4\pi/10$ times the number of ampere-turns gives the magnetomotive force. However, since $4\pi/10$ is a pure number without units, NI itself might be considered to be the magnetomotive force, and, indeed, ampere-turns is the customary engineering unit.

When one erg of work is done in carrying a unit pole once around a magnetic circuit, the magnetomotive force is said to be one gilbert. Thus the expression

$$\int H\,dl = \frac{4\pi NI}{10}$$

gives the number of gilberts and we see that the number of ampere-turns must be multiplied by $4\pi/10$ (approx. 1.26) in order to obtain the number of gilberts. The ampere-turn is, therefore, a slightly larger unit than the gilbert.

Considering a portion of the circuit of length l, we may call the work done in carrying a unit pole over its length the *magnetic potential difference*. Thus, in Eq. 287, H_1l_1 is the number of gilberts of magnetic potential difference across l_1 and, by Eq. 288, $(10/4\pi)H_1l_1$ gives the number of ampere-turns, out of the total number NI, which are necessary to establish that magnetic potential difference. Calling N_1 the turns corresponding to l_1, then we may write

$$N_1I = \frac{10}{4\pi}H_1l_1.$$

Since $(4\pi/10)N_1I$ gives the number of gilberts, this equation tells us then that H_1 in gauss is equivalent to $4\pi N_1I/10\,l_1$ gilberts per cm. Corresponding to the statement at the end of the previous paragraph, we state that the ampere-turn per cm. is, therefore, a slightly larger unit than the gauss (or gilbert per cm.).

We have previously expressed H in gauss and represented it graphically by a number of lines per cm². We have also pictured B by lines of induction per cm². However, since the dimensions of B and H are not the same, they should not be stated in the same

units. Therefore the general engineering custom is to express H in gilberts per cm. or ampere-turns per cm. and B in lines per cm^2. Expressing H in gilberts per cm. is similar to expressing the intensity of an electric field in volts per cm., the volt and the gilbert both being units of potential difference.

184. Analogy with the Electric Circuit. — Using the relations $B = \mu H$ and $\Phi = Ba$, we may reduce Eq. 288 to

$$\frac{4\,\pi NI}{10} = \Phi\left(\frac{l_1}{\mu_1 a_1} + \frac{l_2}{\mu_2 a_2} + \cdots\right).$$

Using the relation $R = \rho l/a$ and calling $1/\rho = c$ the conductivity, we may change Ohm's law from $E = I(R_1 + R_2 + \cdots)$ to

$$E = I\left(\frac{l_1}{c_1 a_1} + \frac{l_2}{c_2 a_2} + \cdots\right).$$

In addition to the similarity in the definition of magnetomotive force in the static circuit and those of e.m.f. in the dynamic electric circuit, we find the following similarity between the equations above. If we choose to think of E as the cause of the current I, we may likewise say that the cause of the flux Φ is the magnetomotive force $4\,\pi NI/10$. Then we see that analogous to the resistance terms there are the $l/\mu a$ terms. These terms are called *reluctances.** We see that there is an analogy between the electrical conductivity and the magnetic permeability. For a given e.m.f. the material of larger conductivity (smaller resistivity) allows a larger current. For a given m.m.f. the material of larger permeability allows a greater flux to exist. The analogy is largely one of external form of the equations; the phenomena are quite different. One of the greatest differences is seen in the analogous terms I and Φ. I is a rate of flow involving motion while Φ is a purely static set of lines of force. In order to keep a current I flowing in a resistance, a continuous supply of energy is necessary. The maintenance of the flux Φ does not require any further energy after it is once established. Furthermore, the conductivity c is independent of the current I, while in the ferromagnetic materials μ depends greatly upon the flux Φ.

185. Methods of Solving Magnetic Circuit Problems. — *Case I. Flux or Flux Density Given.* — Consider a ring such as in Fig. 158 made up of two sections, one of cast iron 25 cm. long

* See footnote, section 128, concerning the units of reluctance.

and 4 cm². in cross section and the other of cast steel 10 cm. long and 2 cm². in cross section. Let it be decided that the total flux through the ring is to be 30,000 lines. How many ampere-turns are required to produce this flux? Eq. 288 gives the answer directly. We must first determine B_1 and B_2 and then obtain the corresponding values of H_1 and H_2. For the cast iron,

$$B_1 = \frac{30{,}000}{4} = 7500 \frac{\text{lines}}{\text{cm}^2.}.$$

From the magnetization curve for cast iron in Fig. 159, we see that the induction of 7500 lines/cm². requires a field of $H_1 = 30$ gauss. Likewise for the cast steel,

$$B_2 = \frac{30{,}000}{2} = 15{,}000 \frac{\text{lines}}{\text{cm}^2.}.$$

From the magnetization curve for cast steel, $H_2 = 60$ gauss. Using Eq. 288, we have

$$\begin{aligned} \frac{4\,\pi NI}{10} &= \int H\,dl = 30 \times 25 + 60 \times 10 \\ &= 750 + 600 \text{ gilberts,} \end{aligned}$$

or

$$\begin{aligned} NI &= (750 + 600) \div 1.257 \\ &= 597 + 478 = 1075 \text{ ampere-turns.} \end{aligned}$$

The magnetomotive force of 750 gilberts or 597 ampere-turns is required to produce the 30,000 lines of induction through the cast iron while 600 gilberts or 478 ampere-turns are required to produce the same flux through the cast steel. If the current is to be limited to 5 amperes, then 215 turns of wire must be wound on the core; or for a current of one ampere, 1075 turns are required. If the windings must be bunched, they should be distributed more on the cast iron than on the cast steel since the drop in magnetic potential along the cast iron is greater than that along the cast steel. Thus less magnetic leakage would result.

Case II. Magnetomotive Force Given. — It often happens that only a certain number of turns may be placed on a core due to its shape and then only a certain maximum current may be used without overheating the insulation. Thus, the value of NI is given. The problem of finding Φ is not quite as simple as the inverse problem which we have just solved.

Consider first a closed wrought-iron ring 50 cm. long with cross section of 10 cm². Let NI be given as 1000 ampere turns. By Eq. 288, we have

$$\frac{4\,\pi \times 1000}{10} = H_1 \times 50.$$

$$H_1 = 25.1 \text{ gauss.}$$

From Fig. 159, we see that $B_1 = 14{,}800$ lines per cm². and so $\Phi = 148{,}000$ lines.

Now consider the case when the above ring is cut through and an air gap of 1 mm. is made. Applying Eq. 288, we get

$$\frac{4\,\pi \times 1000}{10} = 50\,H_1 + \frac{1}{10}\,H_2.$$

Since H_2 is in the air gap and the same number of lines pass through the air gap as pass through the sample, then $H_2 = B_1$, and

$$400\,\pi = 50\,H_1 + \tfrac{1}{10}\,B_1. \quad (289)$$

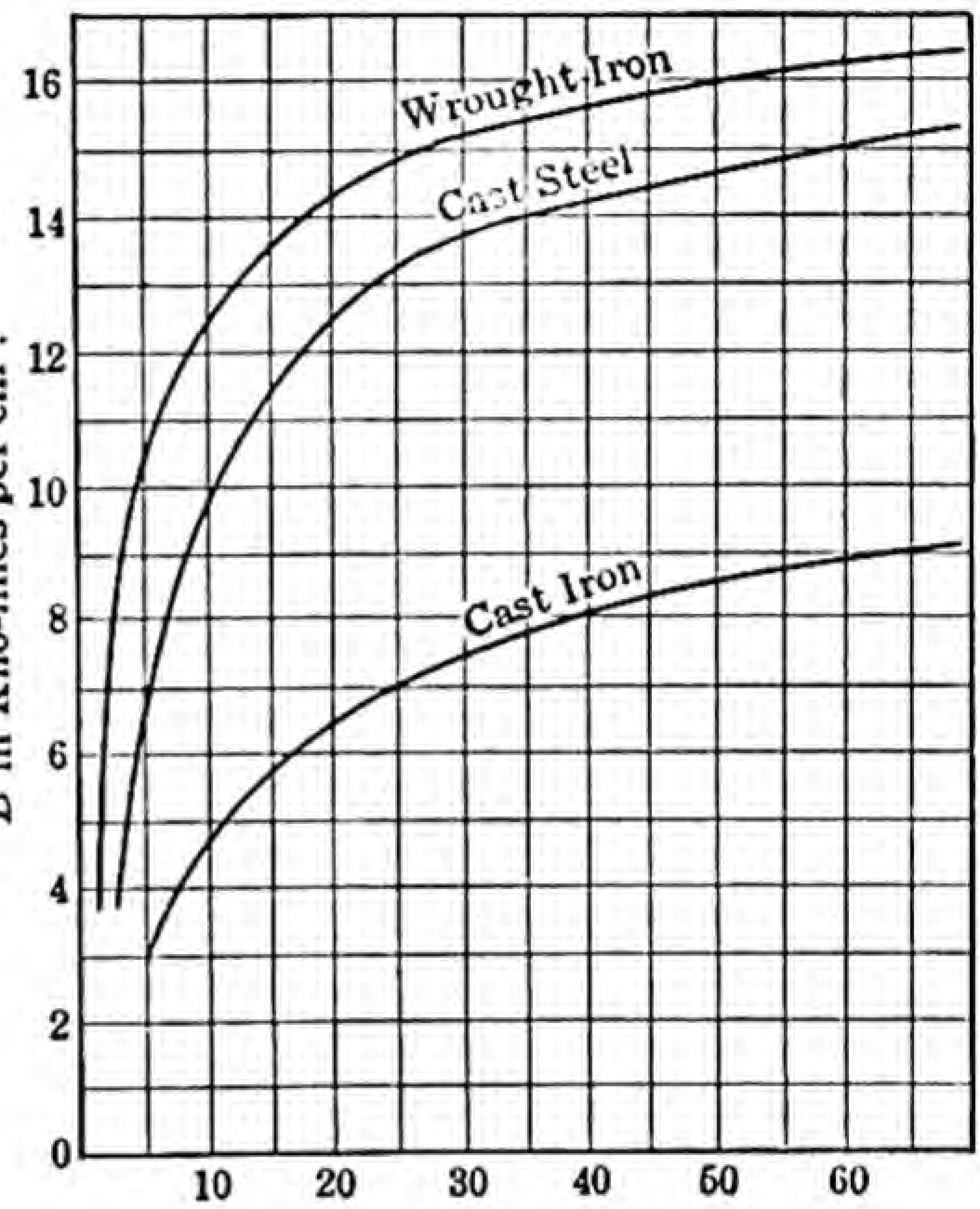


FIG. 159

This equation involves two unknowns. The magnetization curve for wrought iron also involves the same variables: $B_1 = \mu H_1$. The solution of these two simultaneous equations will give the desired values of B and H. Since there is no simple algebraic equation which will represent μ as a function of H, the equations cannot be solved by algebra. The solution is graphically obtained by plotting the straight line relation given in Eq. 289 on Fig. 159. The intersection of this line with the curve for wrought iron gives the induction and field strength existing. The student should plot the line and show that $H = 5$, $B = 10{,}000$, $\mu = 2000$, and $\Phi = 100{,}000$. The tremendous effect of even a very small air gap is seen. From Eq. 288 we find that $(10/4\,\pi)\ 5 \times 50$ or 200 ampere-turns is the magnetic potential difference across the low reluctance path in the iron while the remaining $(10/4\,\pi)\ 10{,}000\ (1/10)$ or 800 ampere-turns is required to maintain the flux through the high-reluctance air gap.

In practice, such a simple case as the one just solved is seldom found. There are usually several different grades of iron or steel in the magnetic circuit. Only a cut-and-try method is then feasible. A guess is made as to how much flux will be produced. By the method of Case I the value of NI which will produce that flux is computed. If this computed value is larger than the given number of ampere-turns, then a smaller value of the flux is assumed and NI is computed again. The process is continued until a value is obtained which is as close as desired.

PROBLEMS

1. In the last paragraph above, solve the problem with and without the air gap for the case of 500 ampere-turns. When NI is 1000 find the effect of doubling the length of the air gap.

2. In the generator shown in Fig. 160, the base is of cast iron and has a section of 200 cm². and an average length of 50 cm. The armature is of cast steel and 30 cm. diameter (approximate length of all lines passing through the armature). Each clearance is 2 mm. The pole pieces are of wrought iron. The average length of the magnetic path through each of them is 15 cm., and their faces each have an area of 100 cm². How many ampere-turns are required to produce 1,500,000 lines across the air gap? Consider that the cross section of the armature is the same as the area of the pole faces.

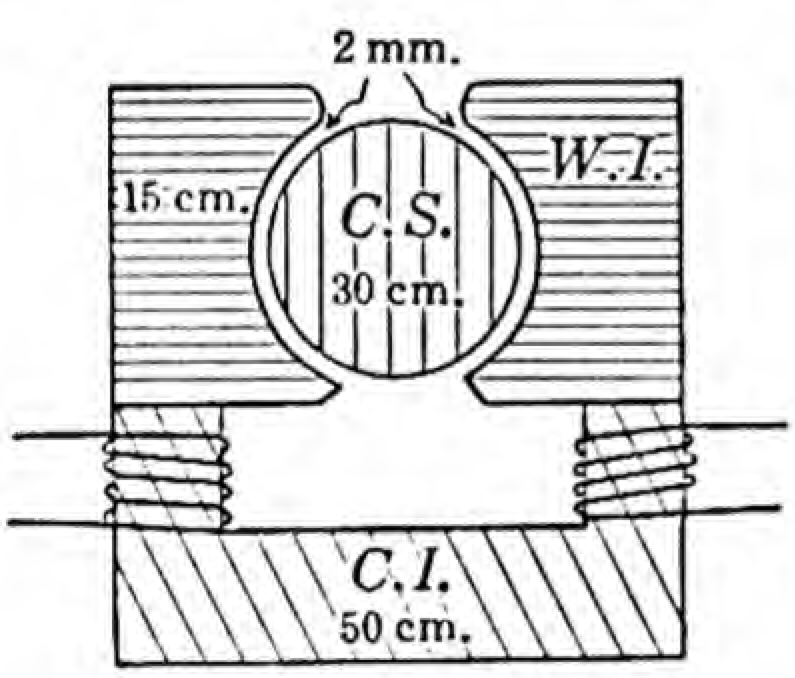


FIG. 160

3. An iron ring with a mean diameter of 20 cm. and section of 10 cm². is magnetized by a coil of 100 turns carrying 5 amp. The permeability of the ring is 1000. Compute the magnetomotive force and express it in two different units. What is the induction in the ring?

4. A ring of 5 cm². in cross section consists of 30 cm. of cast steel, 5 cm. of wrought iron, and 2 mm. of air. What flux will be produced by a 10 ampere current flowing through 300 turns of wire wound on the ring?

CHAPTER V

LIGHT

THE NATURE OF LIGHT

186. Electromagnetic Waves. — The question has been asked, "Does light exist where there is no one to see it?" This is purely a matter of definition. That the light source is emitting energy continually, even in the absence of an observer, can be shown by automatic recording mechanisms designed to record radiant energy. In physics, we define light to be a type of radiant energy such as given off by bodies at high temperatures. In this sense, light is continuously given off by a hot body even though no observer receives the physiological sensation resulting from the absorption of the portion of that energy which falls on the retina.

We observe many properties of light and from these we try to build up a mechanism by which light is transmitted. We know several electrical methods of producing radiations which are not visible. These radiations have many properties which are so similar to visible light waves that we believe them all to be of the same nature. They all have properties of wave motion and their wave-lengths have been measured by various means. Since some of them can be produced and modified by electric and magnetic effects, we believe they are all electromagnetic waves.

We may get some insight into the electromagnetic theory by approaching the subject from the point of view of the long waves of wireless telegraphy. We know how these waves are produced and we know something of the behavior of the alternating electric and magnetic fields associated with them. The long waves of radio telegraphy are radiated from electric charges which oscillate across a spark gap or in a coil. In the case of a spark discharge the terminals of the gap become alternately positive and negative, which means that the electric field oscillates in the plane of the gap. These electric displacements and the magnetic fields which accompany them are propagated outward in all directions

from the source of disturbance. They may set up oscillations in a second circuit properly attuned, the radiation incident upon this second circuit being absorbed. We believe radiant heat and light are of exactly the same nature except that the frequencies are much greater, the oscillators in these cases being either vibrating atoms which are electrically charged or else vibrating electrons. The atoms and electrons may be set into vibration by heat, the long waves (radiant heat) being given off at the lower temperatures, the higher frequencies (light) appearing only at the higher temperatures when the body has become luminous. Vibrations may be set up by other means than heat, as by excitation with ultraviolet light, cathode rays, x-rays, or by radioactive emanations.

The longest waves emitted by electric circuits are several thousand meters, — the shortest are of the order of a millimeter. Rays of nearly one millimeter wave-length are emitted from hot bodies and have been isolated and measured. This region of long invisible waves is called the infra-red region of the spectrum. If the body is at a sufficiently high temperature, there are, in addition, waves sufficiently short to stimulate the nerves of the retina. The wave-lengths of these visible light waves are approximately as follows:

Color	Wave-Lengths
	0.000076 cm.
Red	
	0.000062
Orange	
	0.000061
Yellow	
	0.000056
Green	
	0.000049
Blue	
	0.000042
Violet	
	0.000040

Below 0.00004 cm. (0.4 μ or 4000 Å.) the waves cease to be visible although they are still produced by the same means (incandescent solids or gases or electrically excited gases or vapors). The shortest rays emitted from electric sparks between metal terminals in a vacuum are about 0.01 μ in length. Below this range and overlapping it are waves emitted from x-ray tubes and still further are the extremely short gamma rays given off by radioactive materials.

Possibly, the extremely penetrating cosmic rays may be electromagnetic waves of even shorter wave-length than gamma rays.

187. Frequency. — Consider a vibrating system making n oscillations per second and sending out a continuous set of waves which progress from the source with a velocity v. The first wave given out during a given second of time progresses a distance of v cm. during that second. The n waves given out from the source during that second will be uniformly spaced between the source and the position of the first wave. There will be n points in this space at which the magnitude and direction of the electric field will be the same (also true of the magnetic field). The distance between two such consecutive points, λ, is called the wave-length. There are n of these wave-lengths in the distance v, therefore it is seen that the following relation holds:

$$n\lambda = v.$$

When the waves enter another medium, we find by measurement that the wave-length is changed and that the velocity is changed in the same proportion so that the frequency does not change.

The value of v for visible light as recently determined by Michelson is $(299{,}796 \pm 4) \times 10^5$ cm./sec.; nearly 3×10^{10} cm./sec. or 186,000 miles/sec. So the range of frequencies recognized by the eye is from about 4×10^{14} vibrations per second to 8×10^{14} vibrations per second, the violet light being of the highest frequency.

188. The Photoelectric Effect. — Until recent years no radiation phenomena had been discovered which could not be satisfactorily accounted for by wave theory. We shall discuss one of the several difficulties which have recently presented themselves. The particular one which will be presented is chosen because it emphasizes the importance of the frequency of vibration. According to old wave theory, energy is transferred uniformly over the entire wave front, *i.e.* we expect each atom on the surface of an absorbing body to receive the same amount of energy each second as its neighboring atoms. But we have evidence that this is not the case in the photoelectric effect. *The photoelectric effect is the ejection of electrons from surfaces on which light is incident.* It was discovered by Heinrich Hertz in 1887.

The photoelectric cell consists usually of a glass bulb in which is sealed a metal plate C and a collecting electrode A (Fig. 161).

The plate is kept supplied with electrons by being connected to the negative terminal of a battery. The positive terminal of the battery is connected to the collecting electrode. When light is allowed to fall on the plate, electrons are emitted from it and are attracted to the positively charged rod A. A continuous flow of electrons results and the magnitude of the current is read on the galvanometer G. Careful experiments have shown that the current is directly proportional to the light intensity. The slope of this current-intensity curve depends upon the surface of C. Pure alkali metals produce sensitive cells. If the surface is coated with a thin layer of the hydride of the metal, it is made more sensitive. The most sensitive cells recently made are of extremely thin layers (only a few molecules thick) of caesium on silver.

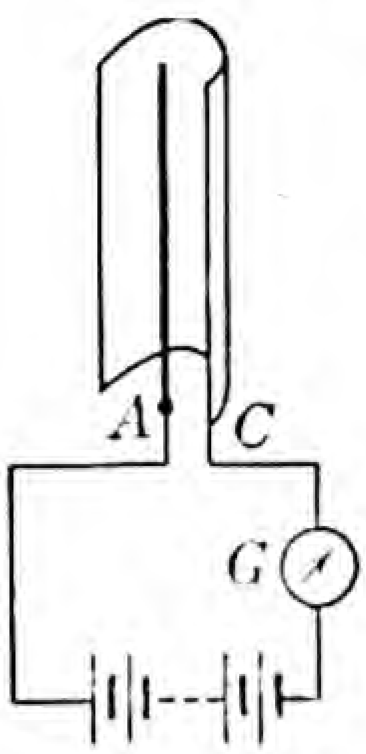


FIG. 161

If the battery is removed and the circuit closed with only the galvanometer across A and C, it is found that electrons still circulate through the circuit. Thus it is seen that the electrons are ejected from the surface with a finite velocity. By applying a variable potential across A and C which opposes the flow of electrons, a potential can be found at which the current is just reduced to zero. Then the electrons of the highest velocity are brought to rest just in front of A and are then repelled back to C. The energy of the fastest electron may be computed from this retarding voltage. When the energy of the electrons of highest velocity is found for light of different frequency and intensity, surprising facts are found. The velocity of emission for a given frequency depends not at all on the intensity of the light. When the light intensity is increased, the current increases, *i.e.* more electrons are emitted, but none of them have larger velocities than the fastest of those under the weak illumination.

When light of too low a frequency (long wave-length) is used to illuminate the cell, no photo-electrons are emitted no matter how intense the light is. As the frequency increases (decreasing wave-lengths), there is found to be a definite frequency, ν_0, for a given material, at which electrons are emitted. The energy of the fastest electrons increases directly with the frequency difference $(\nu - \nu_0)$ as shown in Fig. 162. The threshold frequency (also called the long wave-length limit)

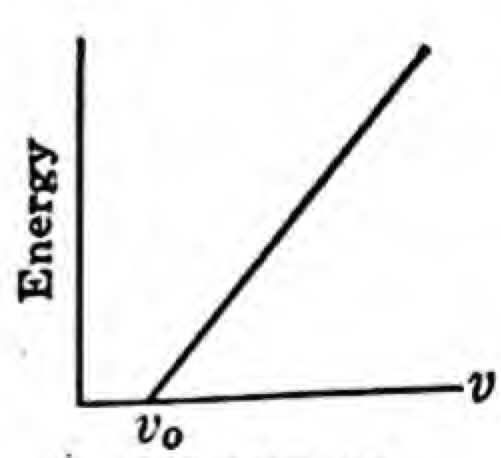


FIG. 162

at which the photoelectric effect begins differs according to the material of the cathode surface.

According to the wave theory every atom on the surface (and probably the next several layers of atoms) should receive equal portions of the incident energy and after absorbing enough energy a photoelectron should be ejected from each atom with as much (or less) energy as experimentally found in Fig. 162. But calculations from the number of electrons emitted per second (determined by measuring the current) by a given very weak source of measured energy flow show that if all atoms should receive energy simultaneously and equally, a time of many hours would have to elapse before the first electron could be emitted. But measurement shows that a flow of electricity starts from the cathode in something less than 10^{-8} second.

This photoelectric paradox and other phenomena have led to the "quantum" hypothesis that in some way the energy in wave motion is not uniformly distributed over the wave front but occurs in bunches or "quanta." The amount of energy in each quantum must be directly proportional to the frequency. So as soon as the light is incident on the cathode, here and there an atom receives a quantum of energy and an electron is ejected with energy as shown in Fig. 162. The equation of the line in the figure is of the form $y = mx - b$ and therefore

$$\frac{1}{2} mv^2 = h\nu - h\nu_0, \tag{290}$$

where h is a constant. Of course, some of the electrons are interfered with and escape with less velocity but none exceed that given in this formula. Eq. 290 was first formulated by Einstein in 1905. No accurate data were available for testing the relation until 1916, when Millikan's precise measurements proved it valid throughout the visible range. Later its accuracy was proved throughout the x-ray region. The value of h is 6.55×10^{-27} erg. sec.

Although modifications of the wave theory have to be made when dealing with the interaction of light with single atoms and electrons, the effects on the average still obey the wave theory accurately.

Luminous Intensity, Illumination, and Brightness

189. Luminous Intensity and Light Flux. — From every luminous source electromagnetic energy radiates at a certain rate and this rate may vary with the direction from the source. The

amount of light energy radiating per second through a unit solid angle in a given direction is called the luminous intensity of the source in that direction. The candle was one of the most common sources of illumination at the time when the comparison of light intensities became a scientific problem, so naturally it was chosen as the standard source. The English Standard Candle was specified to be a spermaceti candle made so as to burn 120 grains per minute with a flame height of 45 millimeters. *The rate at which energy radiates from a standard candle through one steradian in a horizontal direction is called the candle.* We say that a source has a luminous intensity of 5 candles in a certain direction if it radiates energy through one steradian in that direction at five times the rate which a standard candle radiates in a horizontal direction. The terms *candle power* and *luminous intensity* are often used interchangeably because the candle is the most customary practical unit of measurement.

It is very difficult to reproduce the standard candle. The intensity of such a source varies greatly with the size and height of the wick. Many lamps have been devised to replace the standard candle but none have a high degree of reproducibility. Probably the best of these lamps is one designed to burn pentane vapor. The size of the burner openings, the humidity of the air draught, the air pressure, and the purity of the pentane are all very important in keeping the intensity of the light constant. The conditions have been precisely stated for operating the pentane lamp so that it will have a luminous intensity of 10 candles. Because it was found that the light intensity from properly aged incandescent tungsten filaments was far more easily and accurately reproducible than any other source known, an international agreement was made in 1909 upon the relative and absolute candle power of a group of tungsten lamps and these lamps are now accepted as International Standards. They are deposited at the Bureau of Standards at Washington and are used only rarely to calibrate secondary standards for distribution to various countries. The absolute candle powers of these lamps are not definitely known, but they agree closely with the pentane standard. The relative candle powers of the different lamps in the group have been determined to the highest degree attainable, an accuracy far in excess of that of the reproducibility of the pentane standard.

It is convenient at times to consider the case where energy radi-

ates equally in all directions from a source having no finite size. Such a source is called *a point source.* Thus a point source of one candle will radiate through a steradian in any direction as much energy per second as does a standard candle in the horizontal direction and hence radiates more total energy every second than does the whole standard candle.

Many times we wish to consider the average of the rates of emission of energy in all directions about a source. For this purpose we say that a lamp has a *mean spherical candle power* of one candle when the total energy emitted from it each second is the same as would be emitted from a point source of one candle power. As can be seen from the previous paragraph, the mean spherical candle power of the standard candle is less than one candle since its maximum luminous intensity is in the horizontal direction.

At times we wish to consider a certain rate of flow of light energy without reference to the source. Nevertheless the practical unit of this rate is got by referring to the standard candle. *The amount of light energy which radiates per second through one steradian from a point source of one candle power is called the lumen of light flux.* So the light flux from a point source of one candle would be 4π lumens. Although the candle and the lumen have the same dimensions, namely, energy flow per second, the candle power refers definitely to the rate of emission from a definite source while the lumen may refer to a rate of flow of energy without any reference to where or from what it has originated. Their magnitudes are different because of the very definitions. We may write the following equation:

$$I\left(\begin{matrix}\text{the luminous intensity}\\ \text{in candles}\end{matrix}\right) \times 4\pi = F\left(\begin{matrix}\text{the flux}\\ \text{in lumens}\end{matrix}\right). \qquad (291)$$

190. The Inverse Square Law and Illumination. — If we choose two spherical surfaces of radii r_1 and r_2 about a source at S of I candles (Fig. 163) and we choose the radii sufficiently large compared to the dimensions of the source, then the same energy passes per second through the surfaces AA' and BB' which subtend the same solid angle at the source. Since

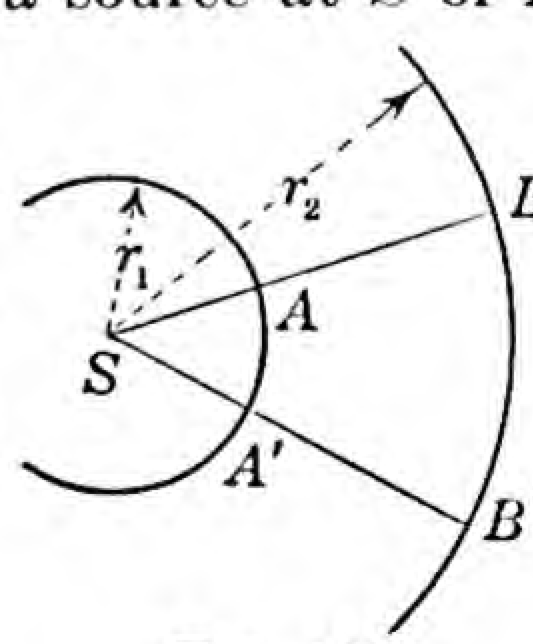


FIG. 163

$$\frac{\text{Area } BB'}{\text{Area } AA'} = \frac{r_2^2}{r_1^2},$$

then the amount of energy passing through a unit area of BB' is smaller than that through a unit area of AA', in the inverse ratio of the squares of the two radii. This inverse square law of radiation is a property of the three-dimensional space in which we live. We call the *amount of energy falling per second on a unit area of a surface, the illumination of the surface.* So, as a property of our three-dimensional universe, the illumination of a surface varies inversely as the square of the distance of the surface from a point source. It is obvious that the amount of energy falling per second on a unit area of any surface such as BB' is directly proportional to the amount of energy which is leaving the source each second. We, therefore, can state that the illumination of a surface is directly proportional to the luminous intensity of the source and inversely proportional to the square of the distance of the area from the source. The letter E is used to denote illumination. Hence

$$E = k\frac{I}{d^2}.$$

The following units of E are chosen so that k is unity. *The illumination of the surface all parts of which are one foot from a point source of one candle is called one foot-candle.* A square foot of area so placed subtends at the source a solid angle of one steradian, so *one foot-candle is equivalent to one lumen of light flux per square foot.* Several other units of illumination are commonly used. The illumination of a surface one meter from a standard candle is called *one candle-meter or one lux;* — evidently the same as one lumen per square meter. The illumination of a surface one centimeter from a standard candle is called one *phot*, which is a large unit and is equivalent to one lumen per cm². or 10,000 lux.

We therefore write

$$E \text{ (illumination)} = \frac{I \text{ (luminous intensity)}}{d^2}. \qquad (292)$$

When I is in candles and d is in ft., E is in ft.-candles. When I is in candles and d is in meters, E is in candle-meters. When I is in candles and d is in cm., E is in phots.

191. Brightness. — So far, we have dealt with point sources of illumination or sources whose dimensions are small compared to the distances considered. Most practical sources cover large areas, so we are forced to consider the *brightness, or the rate at*

which luminous energy leaves each unit area of a source. This idea is of great practical importance. Suppose that a room is illuminated by a 250 watt frosted lamp. Then consider the same lamp inside of a large milk-glass bowl. Neglecting the absorption of the bowl (or increasing the wattage of the lamp to compensate for it), the illumination of all the walls of the room is the same in both cases. However, the lamp by itself emits so much light from each unit area that it will be dazzling to the eye, while the large bowl will have the same light flux spread out over such a large area that it will not impress the observer as being intolerably bright. We may talk of the brightness of a diffusely reflecting surface in the same manner. Some of the units of brightness are: a candle per square inch, a candle per square cm., and one lumen per square cm. (called the Lambert).

192. Lambert's Law. The Flux Emitted from a Flat Luminous Surface. — Suppose a smooth flat surface to be radiating light uniformly over its entire surface. It is found by experiment that *the brightness of the surface is the same no matter from what direction it is observed. This is Lambert's law.* Let a surface of area a (Fig. 164) be the uniformly luminous surface and let the brightness of the surface, as observed from A, be B candles per unit area. The total candle power of the surface as measured at B is Ba candles. Lambert's law states that when observed from C, at an angle θ from the normal, the brightness still appears to be B. At C the light flux appears just as though the emitting area were only $a \cos\theta$ and since the brightness of this projected area appears to be due to B candles per unit area, the total candle power from the whole surface appears to be $Ba \cos\theta$. Thus the effective candle power and hence the emitted light flux per unit solid angle decreases as the cosine of the angle from the normal to the surface. So, as a consequence of Lambert's law, we deduce that if E is the illumination at A, the illumination at C (A and C equidistant from the surface) is $E \cos\theta$.

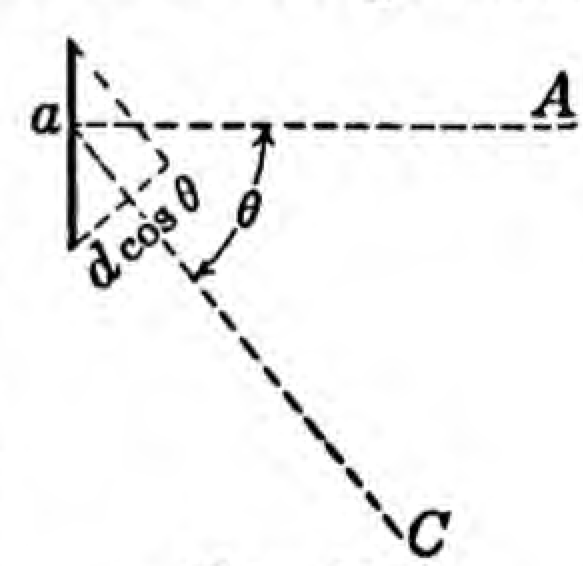


Fig. 164

Since the single surface considered can radiate through only a hemisphere and further since the illumination decreases to zero at the edges of the hemisphere, it is apparent that instead of 4π lumens, not even as much as 2π lumens of light flux can be emitted by each candle of the surface. We shall now compute this total

light flux. Consider a luminous surface of area a (Fig. 165) and a hemispherical surface about it as shown. Let the radius r be large compared to the dimensions of the luminous surface so that any point on the sphere is equidistant from all parts of the surface. Since the light flux per steradian decreases as the angle θ from the normal increases, we will consider the light flux passing through an infinitesimal ring on the sphere, as shown in the figure, all portions of which make the same angle θ with the normal.

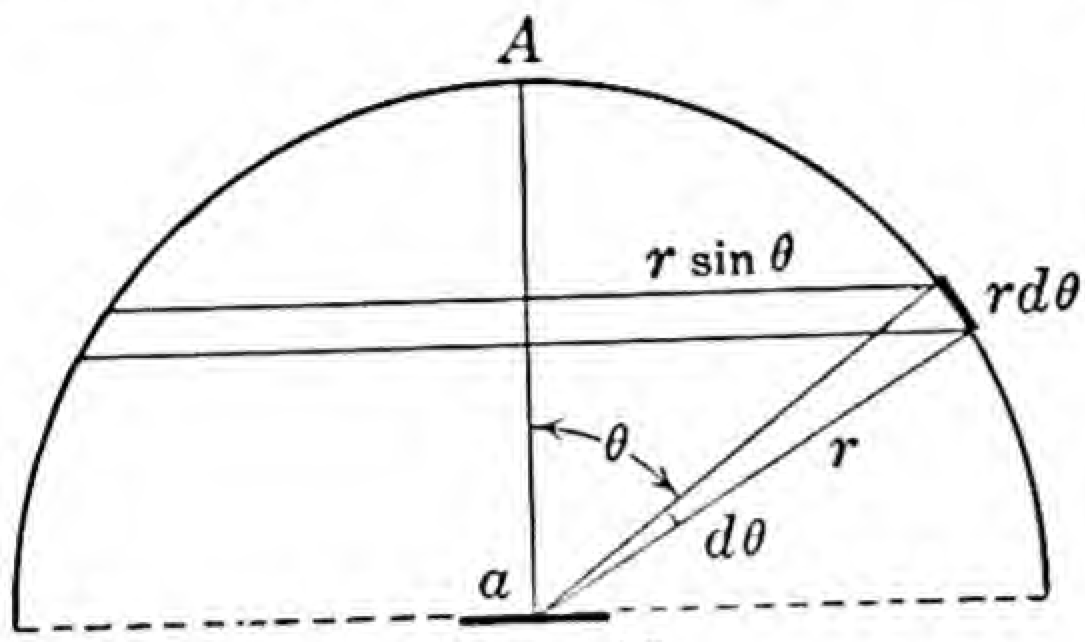


FIG. 165

Let the width of the strip subtend the plane angle $d\theta$. Then the width is $r\,d\theta$. The radius of the ring is $r \sin \theta$ and therefore the area of the strip is $2\,\pi r^2 \sin \theta\, d\theta$. The number of steradians subtended by the area is found by dividing the area by r^2. Hence the solid angle is $2\,\pi \sin \theta\, d\theta$ steradians. As viewed from any portion of the ring, the candle power of the luminous area is $Ba \cos \theta$, and so $Ba \cos \theta$ lumens of light flux pass through each steradian at that angle. Hence, through the area of the ring there pass $Ba \cos \theta\; 2\,\pi \sin \theta\, d\theta$ lumens. In integrating this expression over the hemisphere to find the total flux from the surface, the limits for the plane angle θ must be 0 and $\pi/2$. The total light flux is

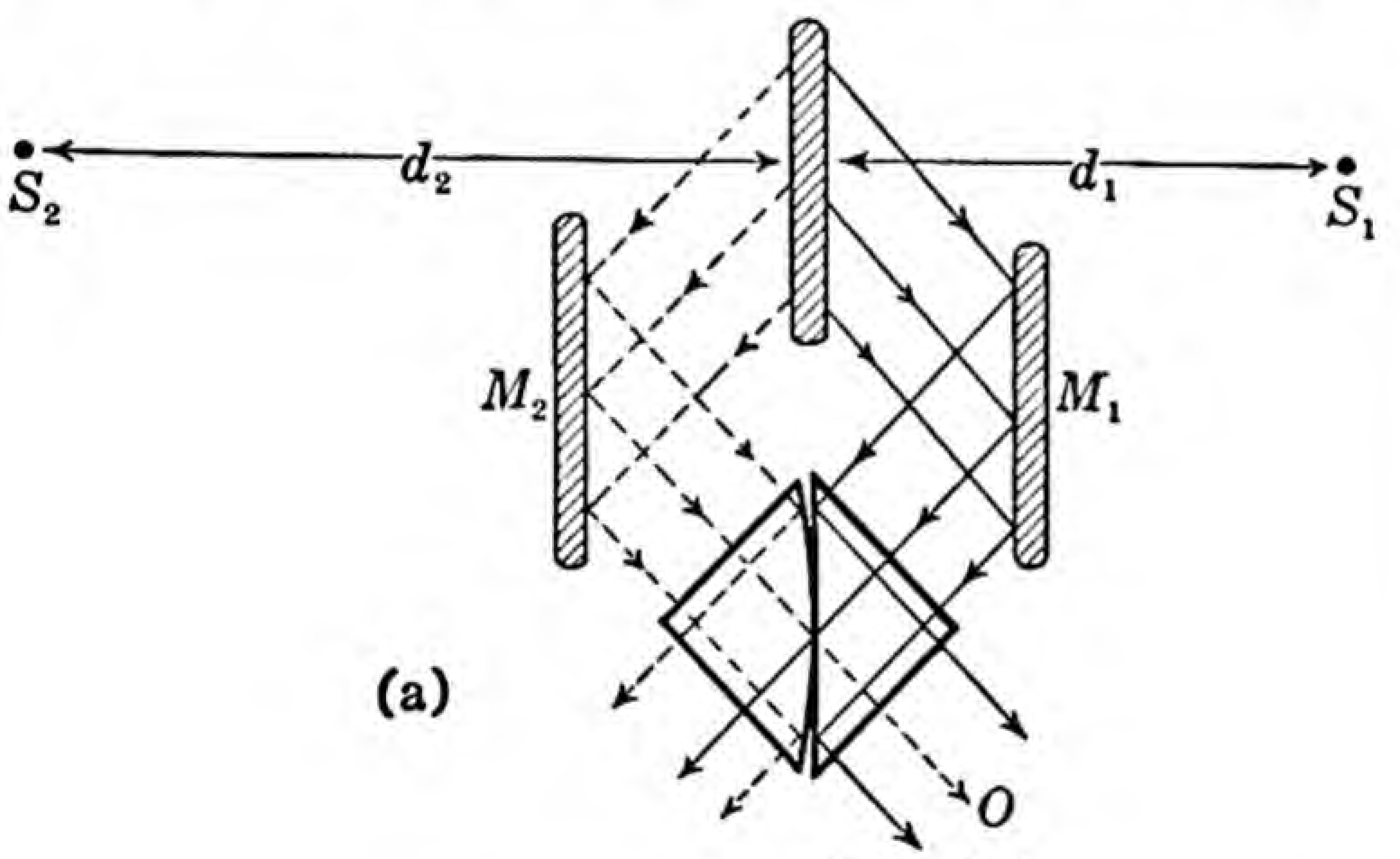
$$F = \pi Ba \int_0^{\frac{\pi}{2}} 2 \cos \theta \sin \theta \, d\theta = \pi Ba \sin^2 \theta \Big]_0^{\frac{\pi}{2}} = \pi Ba. \qquad (293)$$

The total candle power of the surface as viewed from A is Ba. Therefore, we see that from a smooth, plane, luminous surface there are only π lumens per candle.

193. Photometry. — Any device for comparing the luminous intensities of sources is called a photometer. Light from the sources at equal distances may be allowed, in turn, to fall onto the face of a thermopile or into a photoelectric cell. The luminous intensities would be directly proportional to the galvanometer deflections. The thermopile will record all radiation, invisible as well as visible. The photoelectric cell will give comparison of the intensities for only that region of the spectrum to which the emitting surface of the cell is sensitive. If only the visible light is to be compared, then the human eye may be used to observe.

In most usual methods, two adjacent surfaces are illuminated by the two sources respectively and the distances of the sources varied until the illumination of the two surfaces is judged by the observer to be the same. The success of this match depends upon how uniformly the two surfaces are illuminated over their whole area and upon how imperceptible is the division line between the two surfaces. When the dividing edge between the two fields is practically perfect, the eye is extremely sensitive to contrast in brightness. The accuracy is greatly reduced if either or both of the fields of view are not of uniform brightness and of the same color, if the dividing edge is visible, or if the two fields are separated.

One common form of photometer head is that designed by Lummer and Brodhun. It consists of two prisms made as in



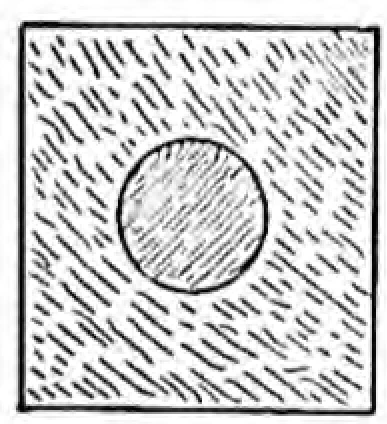


FIG. 166

Fig. 166a. One has a spherical surface with a flattened portion which is polished so smooth that it fits against the other flat prism so perfectly that no reflection occurs at the surface, *i.e.* there is no air between them and they are said to be in "optical contact." The two sources are S_1 and S_2, located at distances d_1 and d_2 respectively from a white screen A. Light from the two sides of the screen which is diffusely reflected in the direction of the mirrors M_1 and M_2, is reflected into the faces of the Lummer-Brodhun cube. The light from M_1 is reflected from the outer edges to the observer at O. The light striking the center portion finds no surface for reflection and passes out the opposite face of the cube. Likewise, light from M_2 passes to O through only the center portion. If the flat face on the spherical surface has been made perfectly so that no corners are nicked off, then the observer sees two fields

of light as in Fig. 166*b* with no visible boundary between them. The sources are moved different distances from A until there is no visible circle, which indicates there is no difference in the light intensities reflected into the two faces of the cube. Assuming the two faces of A equal in reflecting power, then the illumination on its two sides is equal and

$$E = \frac{I_1}{d_1^2} = \frac{I_2}{d_2^2}.$$

In order to compensate for any difference in the surfaces of A, either (1) the sources may be interchanged, a new balance made, and the results averaged, or (2) the whole system of A, M_1, M_2, and the cube may be mounted so as to be turned so that the faces of A are reversed. A slight change in either d_1 or d_2 would balance the photometer again. The average of the two computations for the ratio of d_1/d_2 would be taken as the correct ratio.

Most photometers can be used for comparing the intensity of only those sources that have about the same color. Since the human eye cannot judge when two fields are of the same brightness if they are of different color, the comparison of a carbon lamp with a gas-filled tungsten lamp becomes difficult. Discussion of various special methods which have been developed to photometer lights of different color may be found in books on photometry and in extensive treatises on light.

PROBLEMS

1. The mean spherical candle power of a certain 100 watt lamp is 75 candles. How many lumens of light energy are radiated? What illumination is produced at the following distances: 1 m., 5 cm., 10 ft., 8 in.?

2. In Fig. 166, S_1 is a 50 candle source, $d_1 = 100$ cm., and $d_2 = 125$ cm. when the fields of the viewing prism are matched. Upon reversing the viewing system, $d_1 = 99$ cm. and $d_2 = 123$ cm. What is the candle power of S_2? What is the average illumination of the surfaces of A; expressed in ft. c., c. m., lumens per ft^2., lumens per in^2., and lux?

3. Obtain the relation between the foot candle and the lux. How many candles are equivalent to one lumen?

4. In Fig. 166, $S_1 = 20$ c. and $S_2 = 30$ c. and the distance between them is 200 cm. Find d_1 and d_2. Interpret the meaning of the two solutions.

5. Show that if the line from a source to an area makes an angle θ with the normal to the surface, then the illumination on the surface is $I \cos \theta/d^2$.

6. A small screen is placed directly facing a 75 candle power lamp and 6 ft. from it. A mirror is placed 4 ft. from the line joining the screen and lamp and parallel to the line. The mirror reflects 80 per cent of the incident light. What is the illumination on the screen (consider light coming from both object and image)? (*Ans.* 2.44 ft. c.)

7. A screen is placed 5 ft. from an incandescent strip which is 2.0 cm. long and 3 mm. wide. The illumination on the screen is 16 ft. c. What is the brightness of the strip in candles per cm^2. and lamberts?

8. Two mirrors are placed at right angles to each other. A lamp is placed at some point between the mirrors. Discuss the method of finding the illumination on a screen placed at some other point between the mirrors. What difficulty is encountered if the lamp and screen are placed between two parallel mirrors?

9. A room 15 ft. × 15 ft. × 10 ft. high has a 200 candle power light at the center of the ceiling. What is the illumination on the floor at a corner of the room?

10. A certain square plane white surface diffuses equally in all directions light which is incident upon it. It is placed 5 ft. from a source of 100 candle power and reflects 50 per cent of the incident light. What is the brightness of the surface in candles per ft^2. and in lamberts?

194. Reflection. — The student should review his former work on Huyghens' principle and its applications and also the laws of reflection and imagery in plane mirrors. Practice should be taken in locating graphically the images formed in concave and convex mirrors, remembering that of all the rays leaving a point on an object, four are readily followed: the ray parallel to the axis, the ray passing through the principal focus, the ray striking the mirror where the axis passes through the mirror, and the ray passing through the center of curvature of the mirror. The intersection of these four, or any two of them, after reflection, locates the image of the selected point on the object.

The relations existing between object and image positions for spherical mirrors and lenses are easily found by use of a relation between the sagitta of an arc and the radius of curvature.

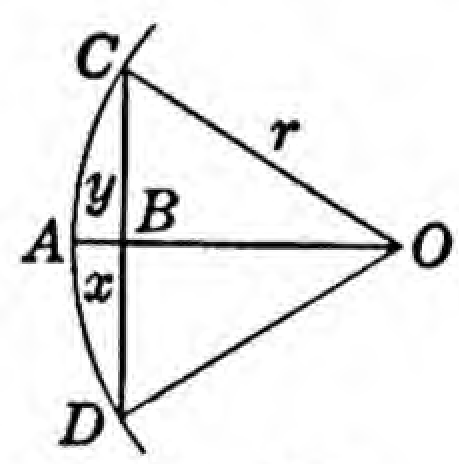


Fig. 167

195. The Sagitta Theorem. — Consider a chord CD $(= 2y)$ of an arc of a circle (Fig. 167). Draw a radius $OA (= r)$ perpendicular to the chord. The portion AB $(= x)$ of the radius between the circle and the chord is called the sagitta (arrow). We shall now obtain a relation involving x, r, and y.

$$r^2 = y^2 + \overline{OB}^2 = y^2 + (r - x)^2 = y^2 + r^2 - 2rx + x^2, \quad x = \frac{y^2}{2r - x}.$$

If the angle COA is very small, then x is negligible in comparison to $2r$. Then

$$x = \frac{y^2}{2r} \text{ when } x \text{ is small.} \qquad (294)$$

The only condition under which distinct images are formed by spherical mirrors is this very same condition.

196. Conjugate Focal Points for a Concave Spherical Mirror. — For every position of an object there is a corresponding image position. These pairs of points are called *conjugate focal points* because if the object is moved from its location to the place where its image is formed, then the image is formed at the former position of the object. In other words, object and image positions are conjugate, or are always interchangeable.

Let an object be located at O, distant p from a concave mirror, of radius r, whose center is located at C (Fig. 168). Consider a wave front from O which has reached the position MAM'. By the time the portion of the wave front at A has advanced to the mirror at P, the portion at M has been reflected to N. The reflected wave front is now NPN' and this wave converges to a point I, distant q from the mirror, and likewise for all succeeding waves so that a real image of O is formed at I. Because we know that all parts of the wave front travel at the same velocity both before and after reflection, we know that the distance $AP = MN$. Now if we limit the size of the mirror so that the angle MOM' is always small, then little difference exists between the distances MN and BD, and likewise between MB and ND. So for small angles subtended by the mirror at the object point,

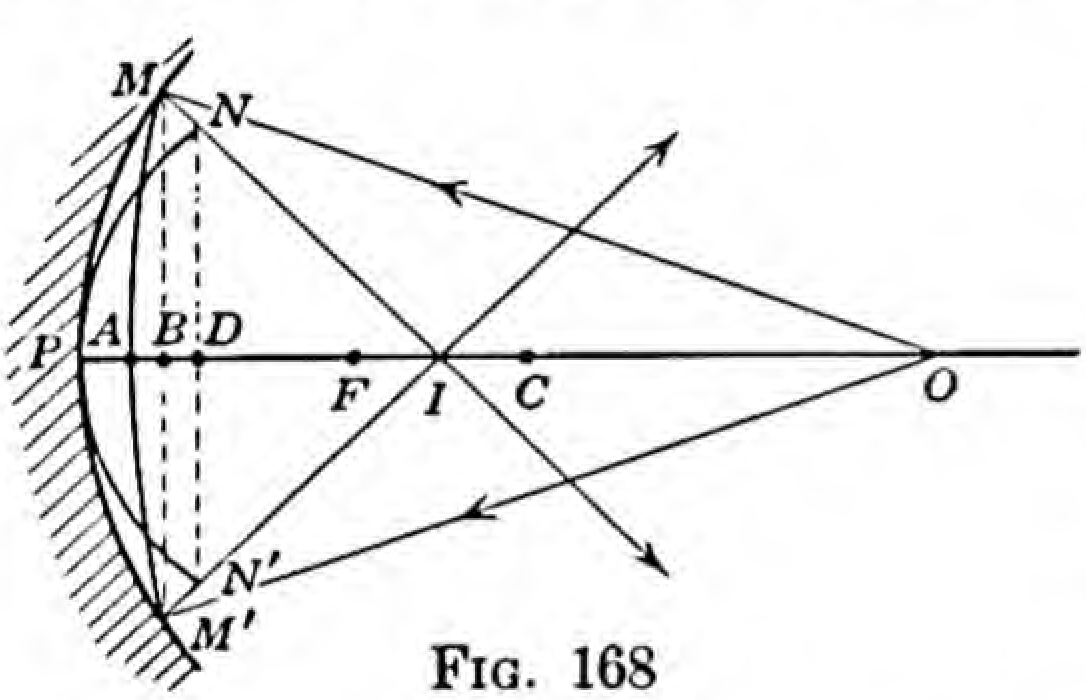


FIG. 168

$$AP = BD. \tag{295}$$

We must now express these distances in terms of the sagittae of the arcs in the figure. The sagitta of MPM' is PB; that of MAM' is AB; and that of NPN' is PD. Expressing Eq. 295 in terms of these sagittae,

$$PB - AB = PD - PB. \tag{296}$$

Calling y the distances MB and ND, and applying Eq. 294 to Eq. 296,

$$\frac{y^2}{2r} - \frac{y^2}{2p} = \frac{y^2}{2q} - \frac{y^2}{2r},$$

or

$$\frac{1}{p} + \frac{1}{q} = \frac{2}{r}, \tag{297}$$

where p is the distance of the object from the mirror, q the image distance, and r the radius of curvature of the mirror. When the object is placed at a very great distance from the mirror, $1/p$ becomes nearly zero and $q = r/2 = f$, say. The point F, distant f from the mirror, is, therefore, the point where all rays parallel to the axis of the mirror cross after reflection or where all plane wave fronts converge. The point F is called the *principal focal point* of the mirror. The plane perpendicular to the axis and passing through F is called the principal focal plane. Then Eq. 297 becomes

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f}. \tag{298}$$

197. The Convex Mirror. — In Fig. 169 is represented a convex mirror NPN' whose center of curvature is at C. A wave from the object O is pictured as MPM' just as it touches the mirror. While the portions M and M' travel to N and N', the portion at P

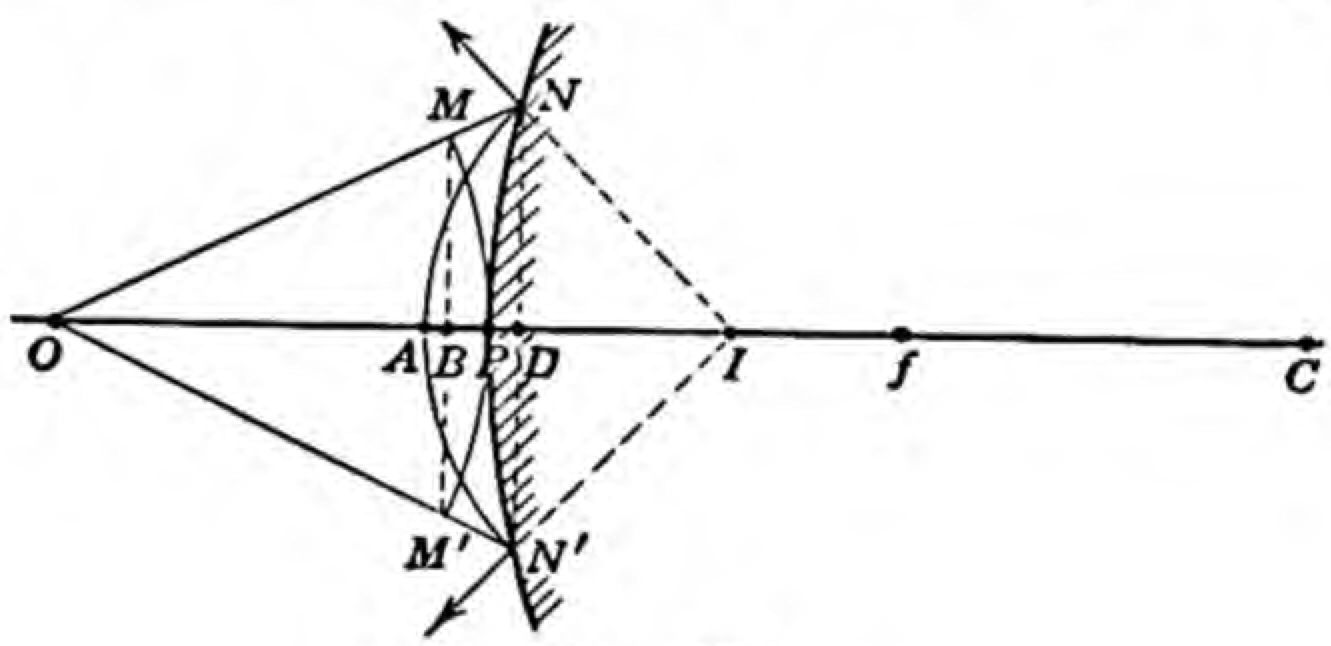


Fig. 169

is reflected and advances a distance equal to MN to point A. The reflected wave front NAN' is divergent and appears to have come from I, the position of the *virtual image*. Again limiting the size of the mirror so that angle NON' is very small, $MN = BD$. In this case, we have

$$AP = BD. \tag{299}$$

These distances are not sagittae of any arcs in the figure, but AD, PD, and BP are. Eq. 299 may be written,

$$AD - PD = BP + PD. \tag{300}$$

As before,

$$\frac{y^2}{2\,q} - \frac{y^2}{2\,r} = \frac{y^2}{2\,p} + \frac{y^2}{2\,r},$$

and

$$\frac{1}{p} + \frac{1}{-q} = \frac{2}{-r} = \frac{1}{-f}. \tag{301}$$

This equation is similar in form to Eq. 298 but the signs of q and f are seen to have changed. This might have been expected, since in Fig. 169 both r and f are measured off on the opposite side of the mirror from that of p.

198. The Rule of Signs for Mirrors. — It is seen, then, that if we take Eq. 298 with all positive signs, we may make it serve for both concave and convex mirrors if we *let p be positive when the incident light rays are divergent and q be positive when the reflected rays are convergent.* This is the usual case. We might expect the rays to diverge from a real object and converge to a real image. *The opposite case causes the sign of the quantity to be negative.* With regard to the sign of f: parallel rays converge to F in case of the concave mirror, sc f is positive. In the convex mirror, parallel rays diverge as if coming from F, so f is negative.

Whenever the object is real, the light rays incident on any mirror must be divergent and p be positive, but if by means of another mirror or lens a set of convergent rays are incident on the mirror, the object distance is taken to be the distance from the mirror to where the rays would converge if the mirror were not there. The object is said to be virtual and p is considered negative.

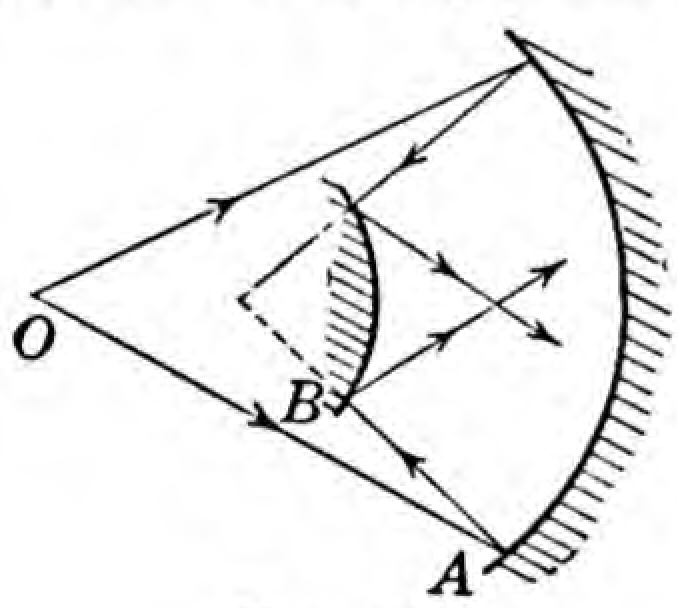


FIG. 170

A typical two mirror example will now be solved. Mirror A (Fig. 170) is 50 cm. from the object at O, and $r = 20$ cm. The mirror B is convex, has a radius of 10 cm., and is 10 cm. from A. For the mirror A,

$$\frac{1}{50} + \frac{1}{q} = \frac{1}{10},$$

so $q = +\ 12.5$ cm. So the rays converge after reflection from A. They converge as if to meet at a point 2.5 cm. behind B. So for mirror B, $p = -\ 2.5$.

$$-\frac{1}{2.5} + \frac{1}{q} = -\frac{1}{5}, \quad q = +\ 5.$$

So the light reflects from B and converges to form a real image half way between the two mirrors. Of course, the rays if not caught on a screen at that point will be incident again on A and a new image formed, etc.

199. The Parabolic Mirror. — A parabolic mirror is formed by revolving a parabola about its axis. Fig. 171 shows the intersection of such a surface of revolution with any plane passing

through the axis. We shall show by use of Huyghens' principle that waves emitted from a point at the focus of the parabola have plane wave fronts after reflection from the mirror.

The parabola $CBAE$ is generated by a series of points which are equidistant from some point O (the focus) and some line DD' (the directrix). Thus $OA = a = AH = g$. $OB = c = BF = e$.

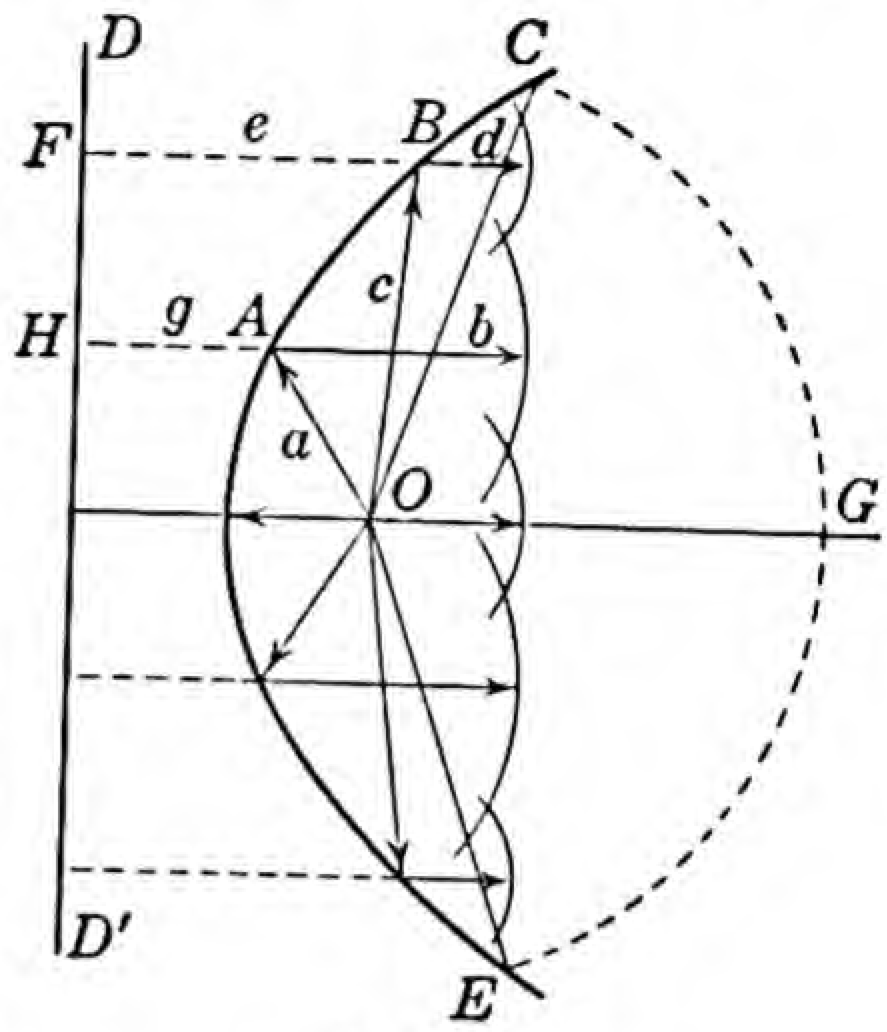


Fig. 171

Now consider a single wave emitted at O. A portion of the wave traveling toward the right is not interfered with by the mirror and its position at a certain time is indicated by the dotted line CGE. In order to follow the reflected portion, we consider each point on the wave front, at the time when that portion touches the mirror, to be a source of a new spherical wave. Thus the portion of the wave which reaches A sends out a spherical wave. While the original wave travels outward a distance of $OC = OG = OE$, the part of the wave from O travels to A and there progresses as a sphere of radius b such that $a + b = OC$. Likewise in the same time interval another portion of the wave which reaches B may be considered as sending out a wave from B which travels a distance d so that $c + d = OC$, and so on for every portion of the wave which strikes the mirror:

$$a + b = c + d = \cdots$$

But by definition of the parabola,

$$a = g,\ c = e, \text{ etc.}$$

Substituting these values in the equation above,

$$g + b = e + d = \cdots$$

But these are the distances from the directrix and so we have proved that the envelope of all the wave fronts reflected from the mirror is a plane passing through C and E. The expanding portion of the wave, CGE, spreads out and the illumination decreases as the square of the distance away and, therefore, rapidly decreases to a very small value, but the light from all the solid angle about the source subtended by the whole mirror is reflected into a plane

wave which must advance forward out of the mirror, its intensity being decreased only by the scattering and absorption of the air.

All search-light mirrors are parabolic. The arc lamp is placed as near as possible to the focal point.

PROBLEMS

1. A concave mirror of radius 50 in. has objects 5 inches high placed at the following distances from the mirror, 10 in., 25 in., 40 in., 50 in., 500 in., ∞. Where are the images formed and how large are they? Are they real or virtual, erect or inverted? Give the corresponding answers for a convex mirror of the same radius.

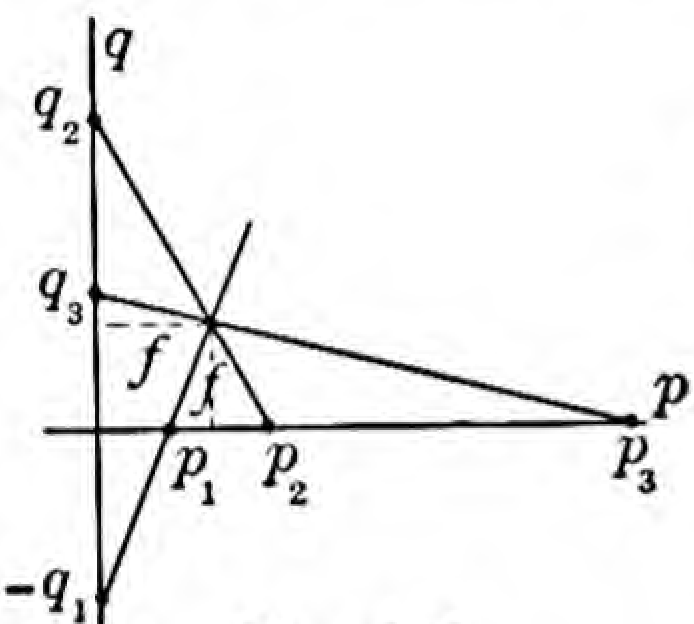


FIG. 172

2. Show that the values of p and q for a concave mirror may be found graphically from Fig. 172, by drawing a line through the object position on the x axis and the point (f, f), the intersection on the y axis giving the image position. Study the motion of image position, q, as the object is moved from the mirror out to infinity. How could the figure be used for converging light where the object is virtual? Find how the figure must be modified so as to apply to a convex mirror.

3. A small source of light is placed halfway between two mirrors that are 10 cm. apart. Mirror A is convex and has a radius of curvature of 5 cm. Mirror B is concave and has a focal length of 2 cm. Find the first two images when the light is reflected first from A and then from B and likewise for the light which strikes first on B and is reflected onto A.

4. The planet Jupiter, as observed from the earth, has an angular size which varies from 32″ to 50″ according to its position relative to the earth. What will be the variation in size of its image formed by a concave mirror of 100 ft. radius of curvature?

5. Show that the curve traced by plotting the points given by the values of p and q in Fig. 172 is an equilateral hyperbola with the origin at (f, f).

REFRACTION AND DISPERSION

200. Snell's Law. — The velocity of light is known to depend upon the medium through which it is traveling. A beam of light is retarded in any material medium, the retardation depending upon the material and the wave-length. The bending of a ray of light in passing from air to water is caused by this retardation. Let us consider a plane wave, AB (Fig. 173), striking obliquely

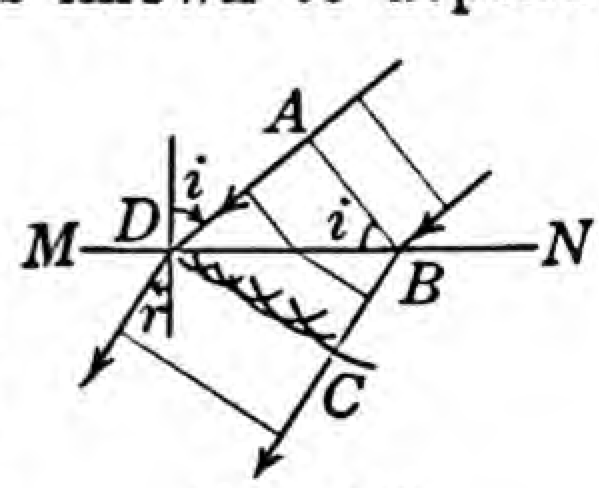


FIG. 173

the surface MN of some transparent surface and let us assume we are dealing with some particular wave-length, λ. The disturbances

set up in the surface MN at the point B do not travel as far in a given time as those originating in the first medium at the same instant at A. Suppose the disturbance from B advances into the medium as far as C while the disturbance from A advances to D. The wave front in the medium is indicated by the line DC. The ratio of the velocities in the two media is given by the ratio AD/BC. That is, $v_1/v_2 = AD/BC$. But $AD/DB = \sin i$ and $BC/DB = \sin r$, where i and r are the angles of incidence and refraction, respectively; that is, the angles between the normal to the surface and the incident and refracted ray, respectively. Therefore,

$$\frac{v_1}{v_2} = \frac{\sin i}{\sin r} = n. \tag{302}$$

This ratio, n, is known as the index of refraction for the two media, for a certain wave-length. If the first medium is a vacuum, the ratio is called the index of refraction for the medium, for the particular wave-length considered. In air or any gas, light travels slower than in a vacuum. The index of refraction of air at one atmosphere of pressure is 1.000293 for the yellow light from a sodium lamp. This value is so nearly unity that hereafter it will be considered unity unless otherwise stated.

The above relation is known as Snell's law, having been first stated by Willebrod Snell in 1621. Snell stated merely that the ratio of the sines of the angles of incidence and refraction is constant for any given two media. Huyghens showed that the ratio of the sines was equal to the ratio of the velocities.

The plane determined by the angle of incidence is called the plane of incidence. It contains the incident ray and the normal to the surface drawn from the point of incidence of the ray upon the surface. A very important fact in refraction is that for isotropic media, *the angle of incidence and angle of refraction lie in the same plane but on opposite sides of the normal,* or in other words, the incident ray, the refracted ray, and the normal to the surface at the point of incidence all lie in the same plane.

It is observed that violet light is more refracted than red light. The cause for this is that the violet light of short wave-length travels more slowly in the medium than does the red light. Hence for the violet, v_2 is smaller and n is larger. The usual variation of n with wave-length is

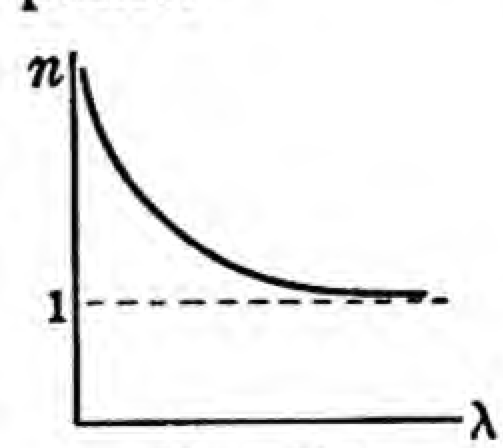


FIG. 174

shown in Fig. 174. The value of n approaches unity as λ increases. If the material has absorption bands in the region considered, then the variation of n is considerably different from that shown.

201. Refraction at the Interface between Two Media. — Let a ray of light AB (Fig. 175) be incident on the boundary between two media of indices of refraction of n_1 and n_2 respectively. Let the velocity of the light in the first medium be v_1, that in the second be v_2, and the velocity in a vacuum be v. By Eq. 302,

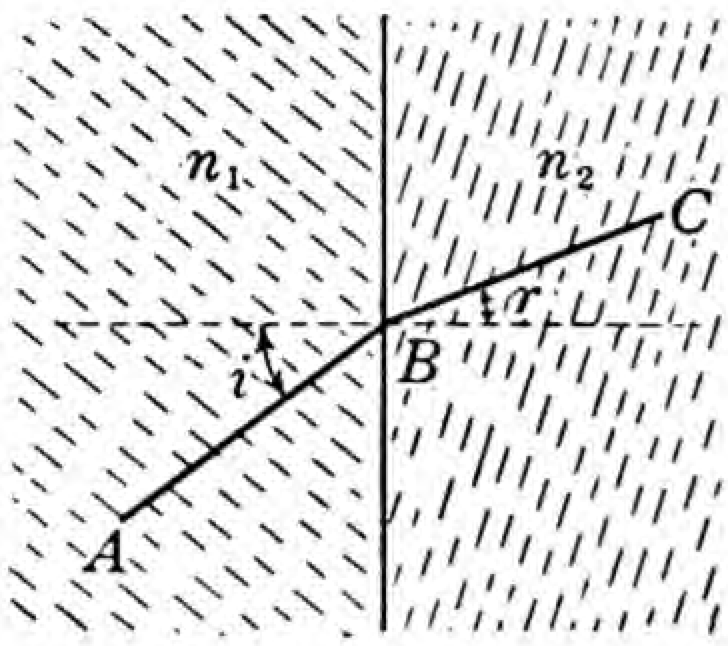


FIG. 175

$$n_1 = \frac{v}{v_1} \quad \text{and} \quad n_2 = \frac{v}{v_2}.$$

If we call $v_1/v_2 = n_{12}$ the relative index of refraction of the two media, then

$$n_{12} = \frac{\sin i}{\sin r} = \frac{v_1}{v_2} = \frac{\frac{v_1}{v}}{\frac{v_2}{v}} = \frac{n_2}{n_1}. \tag{303}$$

Use is made of this formula in computing the deviation of light through lenses and prisms that are formed of two or more pieces of different kinds of glass cemented together.

202. Critical Angle. — If light travels obliquely from a more refractive medium (*i.e.* one in which light travels more slowly)

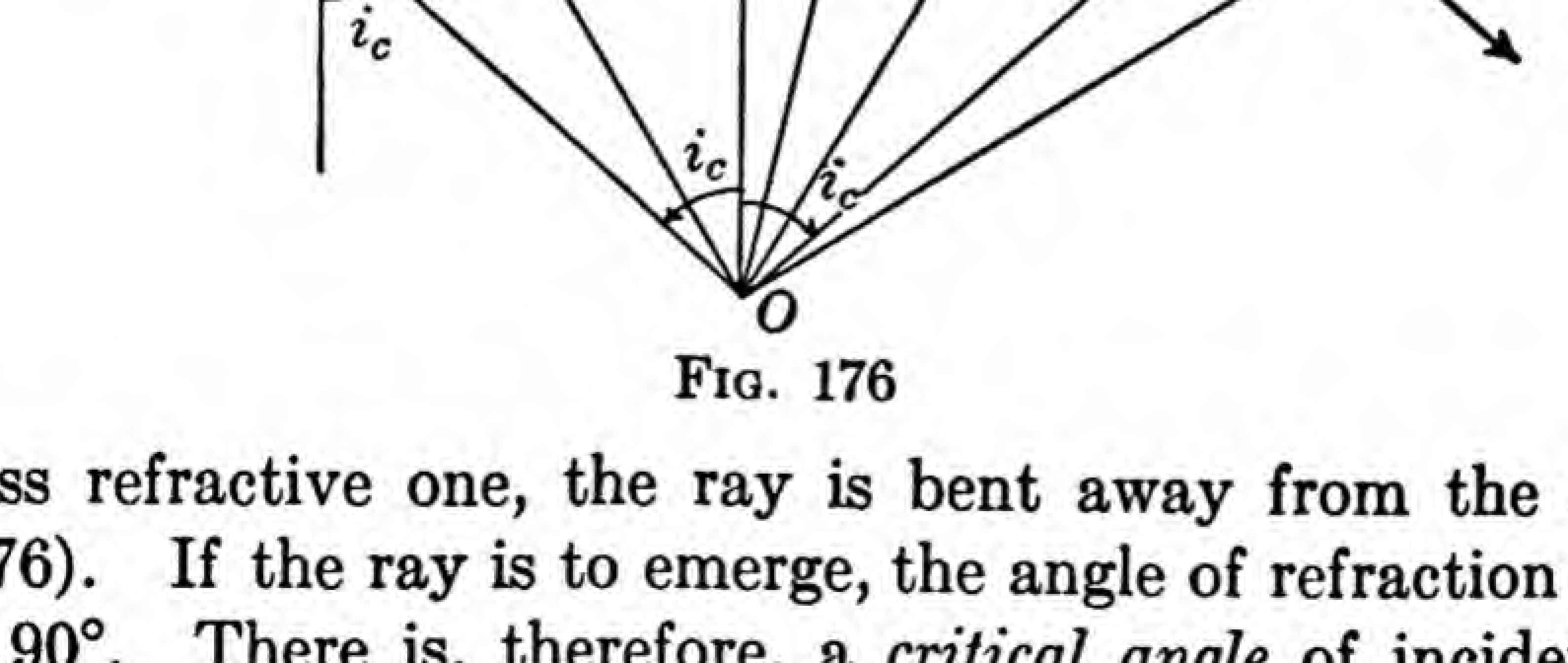


FIG. 176

to a less refractive one, the ray is bent away from the normal (Fig. 176). If the ray is to emerge, the angle of refraction cannot exceed 90°. There is, therefore, a *critical angle* of incidence for the beam within the more refractive medium, for which the angle of refraction is 90° and beyond which a ray is totally reflected. About the point O there is a cone of rays such as OAB making

the critical angle i_c with the normal to the surface. Since the index of refraction is usually taken as the ratio of the velocity in the less refractive medium to that in the more refractive medium and since in this case we are following the ray in the reverse direction, *i.e.* from the denser to the lighter, we must write $n = \sin r/\sin i$ where r is the angle of emergence. The critical angle may be obtained from this equation by placing $r = 90°$. Thus

$$i_c = \sin^{-1}\left(\frac{1}{n}\right).$$

Reflection as well as refraction occurs at the boundary between two media. At normal incidence a maximum amount of light passes into the second medium, only a very small amount being reflected back into the first medium. As the angle of incidence increases the amount of light reflected increases until at the critical angle all is reflected. The intensity of the refracted beam thus diminishes and becomes zero at the critical angle.

203. Apparent Depth. — Let MN (Fig. 177) be the upper surface of a medium and O be the position of an object in the medium. A cone of rays from O, such as COB, diverges further on entering the air above and, to an observer at E, appears to diverge from O'. The larger the cone of rays taken, the more indefinite becomes the point O' for, as seen in Fig. 176, rays approaching the critical angle emerge as if arising from a point only slightly below the surface. However, the pupil of the eye of the observer at E limits the rays which form the retinal image to a very small angle. For such small angles the point O' is quite definite. At B are indicated the angles of incidence and refraction, and angles equal to them are indicated at O' and O. By definition, $\sin r/\sin i = n$. Let d be the depth of O below the surface and d' be the apparent depth or the depth of the virtual image of O. $\sin r = AB/O'B$ and $\sin i = AB/OB$. When i and r are small, $O'B$ is nearly equal to $O'A = d'$ and OB is nearly equal to $OA = d$. So we obtain

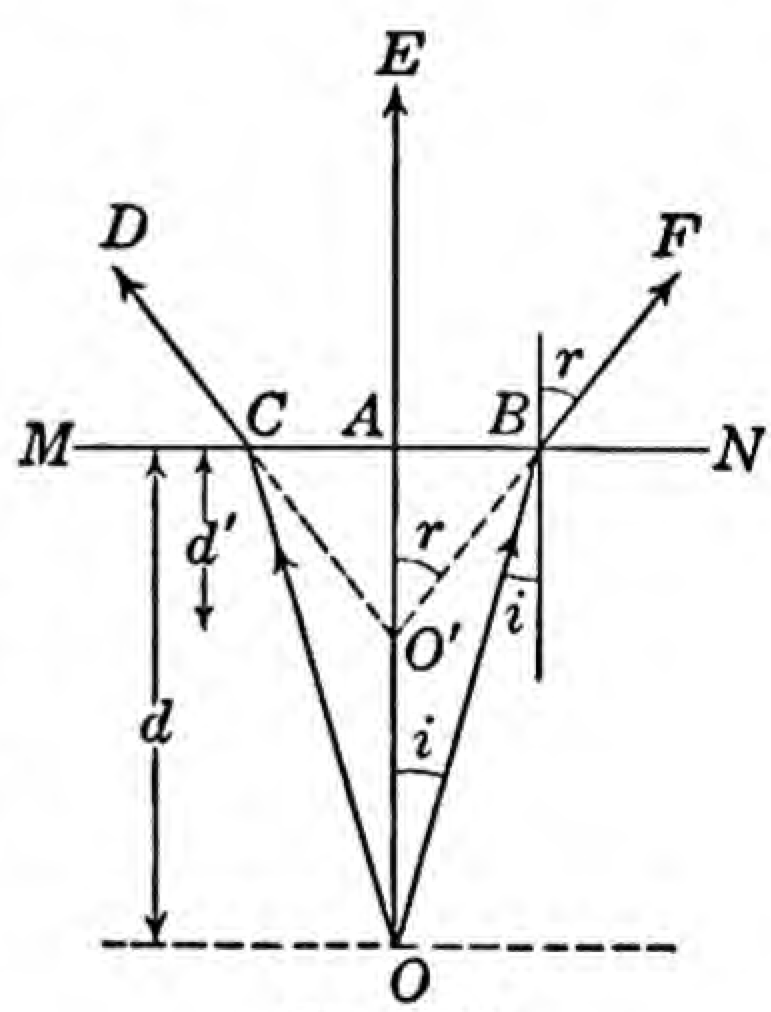


Fig. 177

$$\frac{\sin r}{\sin i} = \frac{OB}{O'B} = \frac{d}{d'} = n. \qquad (304)$$

The value of the index of refraction of a plate of glass or a liquid may be easily determined by measuring the depth of an object in a medium and its apparent depth. A microscope at E is sharply focussed on the object O, in air. The sheet of glass of thickness d is laid on top of the object, or liquid is poured over the object until the surface is a certain height d above O. The microscope must then be raised a distance $d - d'$ in order that O be again seen sharply in focus. The ratio of the two distances d/d' gives the index of refraction. That index of refraction is, of course, the average index of refraction for light of all the wave-lengths visible to the eye and therefore is not very precise.

204. Dispersion. — In § 200 we spoke of the variation of the index of refraction of a substance with the wave-length of the light transmitted through the substance. This variation leads to the separation of white light into the spectral colors. *Dispersion may be defined as the separation of a composite beam of light into its component colors.* If, however, we wish to have definite, measurable values for the amount of separation of the colors, we may say that the dispersion of a substance is the difference of its indices of refraction for any given two wave-lengths of the transmitted light. The variation of the index of refraction of several substances is shown in the table below:

Dispersion Data

	Indices of Refraction					Dispersive Power
Spectral color	Red	Red	Yellow Orange	Blue	Blue	
Fraunhofer designation	B	C	D_1 5896 D_2 5890	F	G	$\dfrac{n_F - n_C}{n_D - 1}$
Wave-length in cm. $\times 10^8$	6867	6563		4861	4308	
A certain flint glass	1.6127	1.6144	1.6193	1.6315	1.6527	0.0276
A certain crown glass	1.5301	1.5311	1.5339	1.5404	1.5509	0.0174
Water (18.7° C.)	1.3310	1.3320	1.3336	1.3380	1.3448	0.018
Carbon bi-sulphide (18.7° C.)	1.6182	1.6219	1.6308	1.6555	1.7020	0.0533

The dispersion of the flint glass may be obtained for any two wave-lengths. Thus,

$$n_G - n_F = 0.0212,$$

or

$$n_F - n_C = 0.0171.$$

205. Refraction and Dispersion in a Plate with Plane Parallel Faces. — It will first be shown that when a beam of light passes through a medium with plane parallel faces, there is no final deviation of the beam, but merely a lateral displacement. A beam of white light, AC (Fig. 178), is incident upon the first surface of the glass. The light of the longest wave-length in the visible red will be refracted along the path CA'. If a ray of the same light were to travel in the direction $A'C$, then it would necessarily emerge from the medium along CA. We see that the beam CA' strikes the opposite face at the same angle that the ray $A'C$ would have struck the first face. Therefore the emergent beam, A_rR, must make with the second face the same angle which the incident ray, AC, made with the first face. Likewise the ray of the shortest wave-length in the visible violet entering along AC will travel along $CA''A_vV$. These two rays and all those of other wave-lengths will be parallel after passing through the plate, but no two of them will be coincident. The same will be true for rays BD and all rays parallel to it. Thus a wave front AB of heterogeneous light will become $A'B'$ (for the longest red ray) and various retarded and rotated wave fronts such as $A''B''$ (for the shortest visible violet). After passage through the plate, there are various wave fronts from A_rB_r for the red to A_vB_v for the violet. In the portion of the emergent beam between A_v and B_r, there is a mixture of all colors and the result will be white light. But just past A_v the light will begin to be deficient in the violet rays. In the region about half way between A_v and A_r, about the lower half of the colors (violet, blue, green) will be missing, and at A_r only the extreme visible red will be present. On the other side of the beam, the light past point B_r will begin to be deficient in red and at B_v will be only the extreme visible violet light. The width of the colored edge of the transmitted beam increases as the angle of incidence increases and as the thickness of the plate increases.

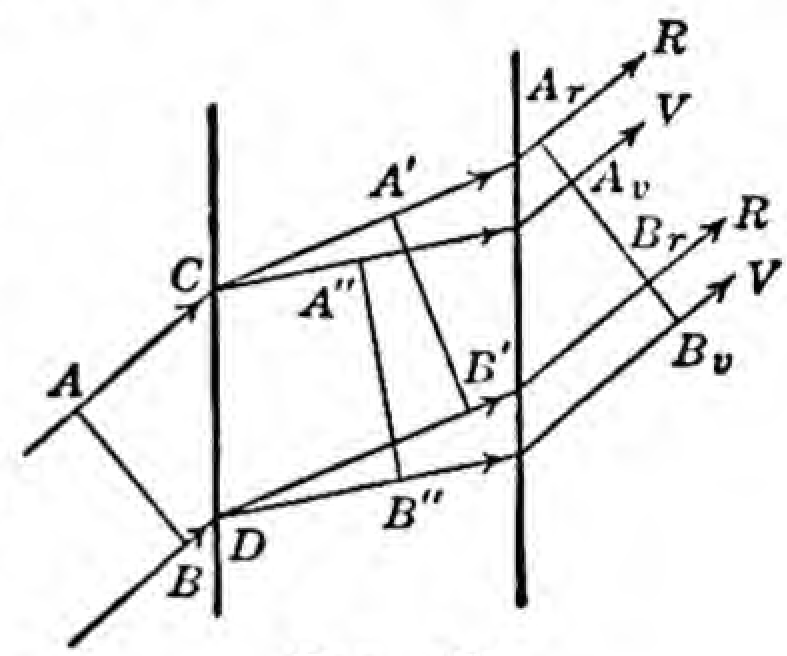


FIG. 178

206. Refraction, Deviation, and Dispersion in a Prism.— If the faces of the plate treated in the above paragraph are not parallel, we have a prism. Then the rays of light in the emergent beam are not parallel to the incident beam, so that both deviation and dispersion result. The deviation and dispersion depend in a complicated way upon the angle of incidence, the angle between the faces of the prism, the frequency of the light, and the index of refraction of the medium.

Let a ray of monochromatic light such as MN (Fig. 179*a*) be incident upon a prism. The angle of deviation D (Fig. 179*b*) is the angle between the incident and emergent beams. When i, the angle of incidence, is changed by rotating the prism (Fig. 179*a*),

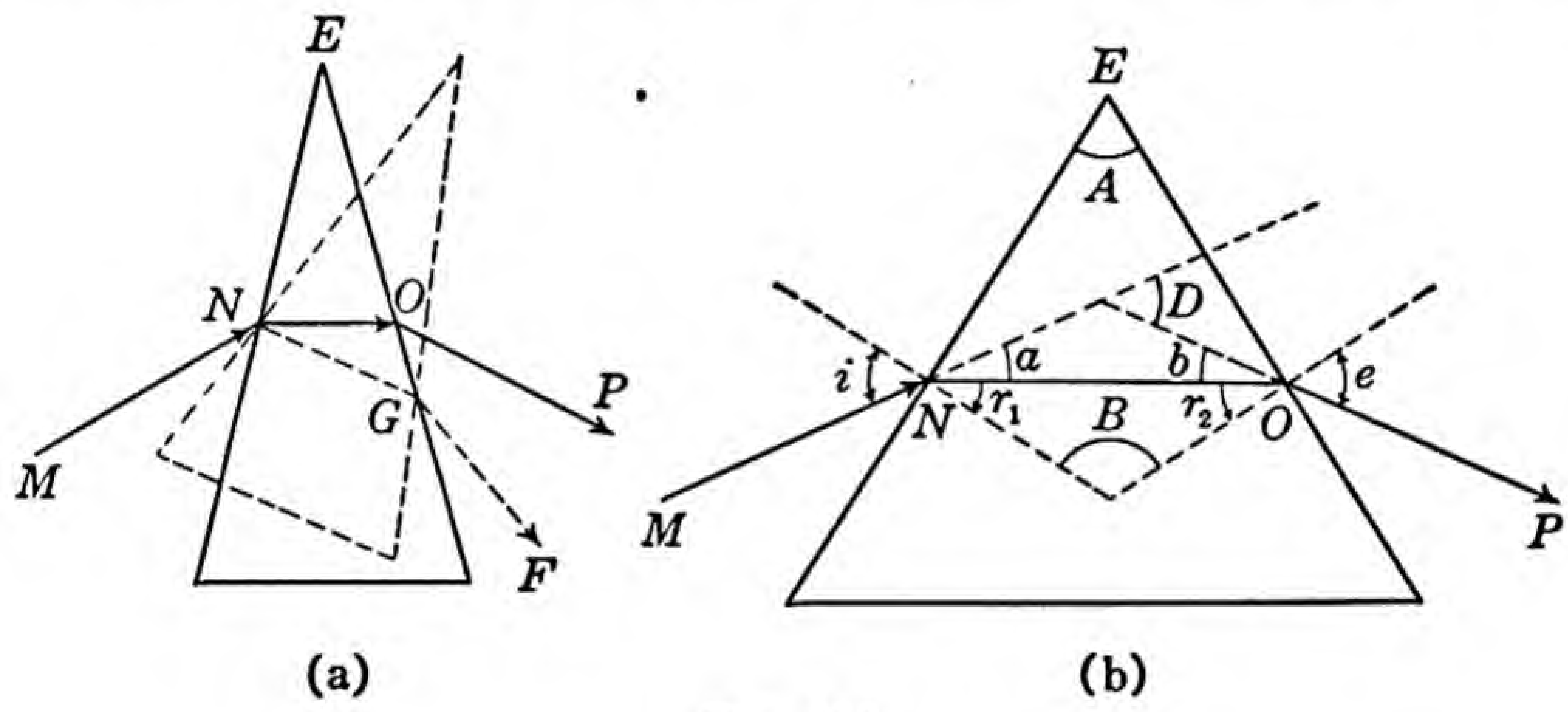


FIG. 179

it is found that the angle of deviation D never decreases below a certain value, which is therefore called the angle of minimum deviation. It is found that, for minimum deviation, the incident and emergent points N and O are equidistant from E, the refracting edge of the prism, and also that the angle of incidence i, in Fig. 179*b*, is equal to the angle of emergence e. No matter which way the prism is rotated, clockwise or counterclockwise, from this position, *i.e.* no matter if the angle of incidence is decreased or increased, the deviation is increased. The dotted lines in Fig. 179*a* show the position of the prism such that the angle of incidence is increased and the angle of deviation, between MN and GF, is increased. Corresponding to each deviated ray such as GF, produced by a clockwise rotation of the prism, there is found a certain position of the prism obtained by a counterclockwise rotation which will produce the same deviation.

In order to prove that at minimum deviation the angles of incidence and emergence are equal, we must first remember the experimental fact that there is only one angle of incidence which will give minimum deviation. Let us assume that the angle of emergence, e, is not equal to the angle of incidence, i. Now consider the ray PO. It follows the path $PONM$ and therefore undergoes the same deviation, D, which is the minimum deviation. But by assumption, the angle of incidence of PO is different from that of MN. We therefore have the case of two rays making different angles of incidence, both giving minimum deviation, but experimentally that is impossible. So we conclude that the rays MN and OP make equal angles with their respective surfaces.

When the prism is set for minimum deviation, the relation between the angle of incidence, the angle of emergence, the angle of deviation, and the index of refraction for light of a given wave-length is quite simple, as shown below.

In Fig. 179*b* are shown the angles of incidence and refraction for a ray at minimum deviation. By inspecting the figure we see that the angles A and B are supplementary, as are likewise angles B and $(r_1 + r_2)$. So,

$$A = r_1 + r_2. \tag{305}$$

Also

$$D = a + b.$$

Then

$$A + D = (a + r_1) + (b + r_2) = i + e. \tag{306}$$

The above relations are true regardless of the angle of incidence, but when minimum deviation is obtained, we have in addition to the above equations,

$$i = e \quad \text{and} \quad r_1 = r_2.$$

In this case, then

$$r_1 = \frac{A}{2} \quad \text{and} \quad i = \frac{A + D}{2}. \tag{307}$$

By definition,

$$n = \frac{\sin i}{\sin r_1} = \frac{\sin \dfrac{A + D}{2}}{\sin \dfrac{A}{2}}. \tag{308}$$

It is seen that this equation gives an easy means of measuring the index of refraction of a substance for light of different colors. A beam of light of a given wave-length is passed through a prism of the desired substance and the prism rotated until the emergent beam is the least deviated. The angle of deviation and the refracting angle of the prism are then measured and n is computed.

The condition for minimum deviation can be rigorously proved as follows:

From Eq. 306, $D = i + e - A$. Also we have $\sin i = n \sin r_1$ and $\sin e = n \sin r_2 = n \sin (A - r_1)$. Hence

$$D = \sin^{-1}[n \sin r_1] + \sin^{-1}[n \sin (A - r_1)] - A.$$

In this expression we have D expressed as a function of only one variable, namely r_1. When D is a minimum or a maximum, $dD/dr_1 = 0$. The student should carry out this differentiation and show that the condition under which the derivative becomes zero is $i = e$. That this condition gives a minimum value for D instead of a maximum may be proved by obtaining the second derivative and seeing that it is positive under the condition $i = e$.

207. The Formation of a Pure Spectrum. — If a point source is placed near a prism as in Fig. 180*a*, the light that is incident upon the prism will be bent very much as in Fig. 178. However, there are two important differences between the two figures: (1) the emergent beams in case of the prism are not parallel to the incident beam but undergo a deviation and (2) since the deviation depends on the angle of incidence, not all the rays will be deviated equally. In general, however, the light received on the screen AB will be predominantly red at one edge of the beam and violet at the other. When a slit is placed in front of the prism so that the image on AB is small, then the white portion of the middle of the beam is quite narrow and the overlapping of the colored regions such as RR and VV is smaller and a better spectrum is formed. We say that the spectrum due to a point source and a slit close to the prism is very "impure" since at every point in the spectrum there is a mixture of light of different wave-lengths.

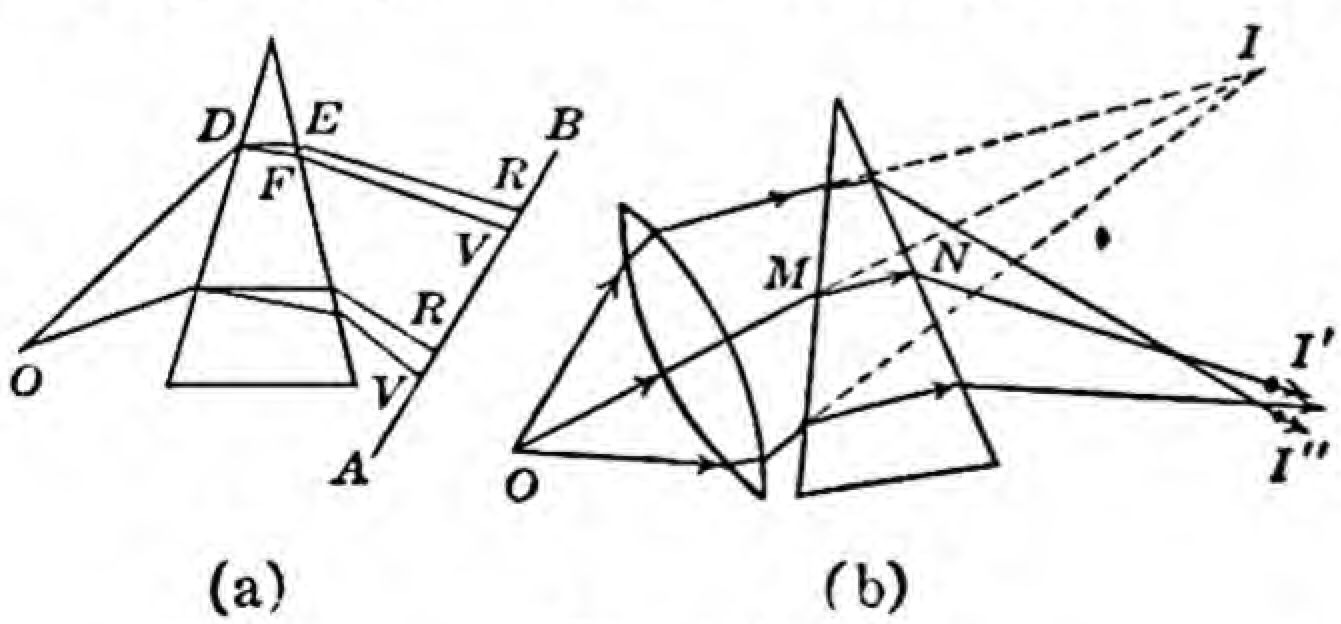


Fig. 180

In order to form a somewhat purer spectrum, a lens may be inserted as in Fig. 180*b*. For light of a certain wave-length the image of a point source O would be formed at I if it were not for the prism. The prism, however, deviates the whole set of rays but does not deviate all the rays through the same angle. If the point I' is

the image of O formed by the narrow bundle of rays about the path OMN which undergo minimum deviation, then the other rays both above and below that bundle will be deviated more, the most deviated forming an image at I''. Then the image of O is spread over the region $I'I''$ for one wave-length. Waves of shorter wave-length will produce other images which will be deviated more from the position I. But it is seen by comparing (a) and (b) of Fig. 180 that the region over which the image for a given wave-length is formed with the aid of the lens is much smaller than without the lens. In other words we have obtained a much purer spectrum.

A still better spectrum is obtained with the use of two lenses as shown in Fig. 181. A slit S is placed at the principal focal point

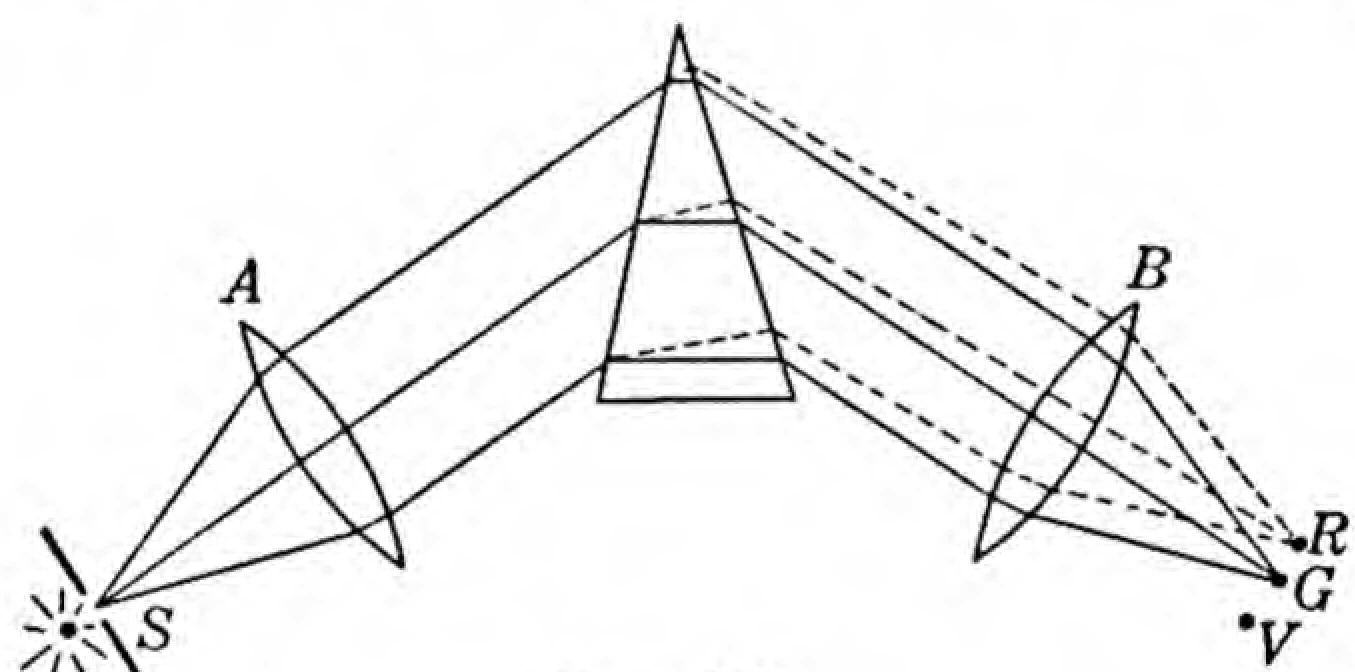


FIG. 181

of the lens A. The light from the source is thus limited to so narrow a beam that the slit becomes practically a "line" source. The length of the slit is perpendicular to the paper. The lens A is constructed so that, as far as possible, light of all wave-lengths emerges from the lens in parallel rays. These rays strike the prism and all of the light rays of one wave-length are deviated the same amount. The heavy lines indicate the rays of a certain wave-length in the green portion of the spectrum, say. Suppose that those rays are brought to a focus by the lens B at a point G. If the focal lengths of both lenses are the same, then at G is an image of the slit of the exact dimensions of the slit. Of course, since the image at G is of finite size, then the image of the slit for a wave-length differing very slightly from that which formed the image at G, may overlap that at G. But with exceedingly narrow slits and good lenses to produce sharply outlined images, a spectrum of exceedingly great purity may be formed. Rays of red light will not be deviated as much as the rays of green light. The dotted lines in the figure show

the paths of rays of a certain wave-length in the red portion of the spectrum and the formation of the image of the slit at R. The blue and violet light will form images on the opposite side of G.

208. Dispersion Produced by Prisms. — For a given angle of incidence of a light ray on a prism face, the angle of the refracted ray may be computed by Eq. 302 for a given wave-length. Then, knowing the angle between the two faces of the prism, the angle of incidence on the second face can be found and another application of Eq. 302 will give the angle between the emergent ray and the normal to the surface. Then the total deviation is easily obtained. Carrying out the same computation for a ray of different wave-length, the amount of angular separation between the rays of two different colors is obtained. For prisms of large angles, no other method is possible, but an approximate value may be easily obtained for prisms of small angle where the incident ray makes the proper angle for minimum deviation.

When the refracting angle of a prism, A, Eq. 308, is small, the deviation D is small and we may write

$$n = \frac{A + D}{A},$$

or

$$D = A(n - 1). \tag{309}$$

Although this equation does not give accurate values for prisms of large angles, it is often used to get an approximate answer.

We have seen in Fig. 180a that when a ray of light OD passes through a prism, the deviation varies with the wave-length. Consider light of three wave-lengths, the C, D, and F spectral lines of Fraunhofer. It is of great practical value to inquire into the relation of D_D to $(D_F - D_C)$ for different glasses, because that will tell us whether different sized prisms of different kinds of glass, made so as to have spectra subtending the same angle, will or will not have some intermediate ray always deviated the same, or vice versa. Using Eq. 309,

$$D_F = A(n_F - 1),$$
$$D_C = A(n_C - 1),$$
$$D_D = A(n_D - 1),$$
$$(D_F - D_C) = (n_F - n_C)A = \frac{n_F - n_C}{n_D - 1} D_D. \tag{310}$$

Now we see that if prisms of different substances are selected with proper angles, A, so that they all give the same deviation D_D for

the D spectral line, then the spectral angles $(D_F - D_C)$ will be different for each prism if $(n_F - n_C)/(n_D - 1)$ is different for each substance.

The table in § 204 shows the variation of this fraction. This fraction might have been evaluated for any three spectral colors whatever, but these three are convenient and are now used throughout the optical industry of the world. By Eq. 310 the fraction

$$\frac{n_F - n_C}{n_D - 1} \text{ is equal to } \frac{D_F - D_C}{D_D},$$

which is the ratio of the angular separation of two colors $(D_F - D_C)$ to the deviation of a chosen ray whose wave-length is about midway between the other chosen wave-lengths. Examination of the table will show that for prisms giving the same average deviation of the spectrum, the spectral angle for the certain flint glass is almost twice that of the crown glass, while that for the carbon bisulphide is nearly twice that of the flint glass. It is appropriate therefore, that the fraction $(n_F - n_C)/(n_D - 1)$ is called *the dispersive power of a medium.*

Since prisms of different kinds of glasses do not produce dispersions proportional to the deviations of the spectrum as a whole, it is possible to combine two or more prisms to produce dispersion without deviation or deviation without dispersion.

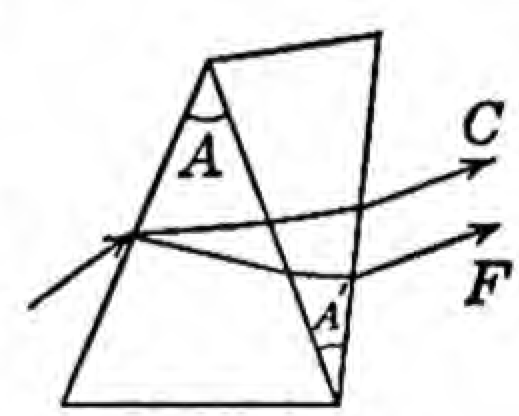


FIG. 182

(1) *Deviation without Dispersion.* — Consider a crown glass prism of angle A which produces a certain deviation and a certain amount of dispersion. A flint glass prism of smaller angle, A', may be made which will produce just as much dispersion without as large a deviation. Take a pair of prisms mounted as shown in Fig. 182. The first prism produces a certain deviation and dispersion as shown. The violet ray is deviated downward more than the red ray by an angle θ, say. The second prism is made of glass of greater dispersive power. Its refracting angle A' is such that the violet rays are deviated upward more than the red ray by the same angle θ. But this second prism produces that same dispersion with a small deviation so that the red and violet rays have undergone a net deviation but emerge parallel to each other. Applying Eq. 309 to this achromatic prism combination, we obtain

for the net deviation of the C and F lines, using primes for the properties of the second prism,

$$D_F = A(n_F - 1) - A'(n_F' - 1)$$
$$= D_C = A(n_C - 1) - A'(n_C' - 1).$$

Then $$A(n_F - n_C) = A'(n_F' - n_C'). \tag{311}$$

By use of this formula, the angle to be used for the second prism may be found corresponding to any angle chosen for the first prism.

It is to be noted that the combination is achromatic for only the two colors selected and all rays of other colors emerge with slightly different angles. By combining three prisms of different dispersive powers, rays of three portions of the spectrum could be made to emerge parallel and, of course, the remaining portions of the spectrum could not deviate from these as much as for the two-prism combination.

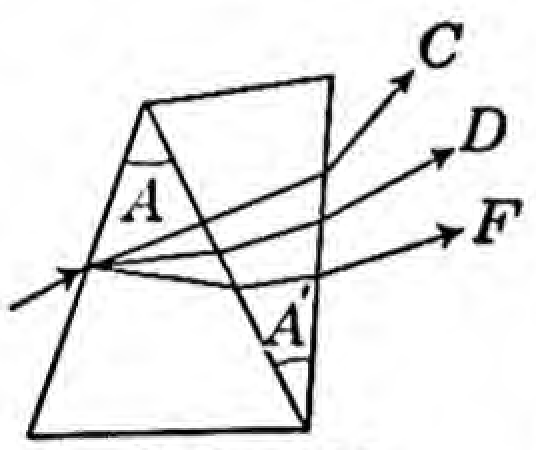


FIG. 183

(2) *Dispersion without Deviation.* — As the angle A' of the second prism is increased (Fig. 183), the rays are bent back more and more towards the incident direction. A certain size of A' can be found where a ray of a certain color is deviated equally by each prism and emerges parallel to the incident ray. Choosing the Fraunhofer D line for such an undeviated ray, we obtain from Eq. 309,

$$D_D = 0 = A(n_D - 1) - A'(n_D' - 1).$$

Then $$A(n_D - 1) = A'(n_D' - 1). \tag{312}$$

Hence for each angle A selected for the first prism an angle A' for the second prism may be found so as to produce zero deviation of any chosen spectral line. Rays of light of shorter and longer wave-lengths such as the C and F lines will be deviated as shown in Fig. 183.

PROBLEMS

1. Prove Eq. 304, using the sagitta method.

2. A ray of white light makes an angle of 60° with a surface of carbon bisulphide. What angular separation have the B and G lines in the liquid?

3. What are the critical angles for water for the B and G lines?

4. A ray of light from an object 10 ft. under water emerges from a surface so that the C line makes a 30° angle with the normal. What angular separation have the emerging C and F lines? Work this problem where the given angle is 80°.

5. In the above problem, how far below the surface of the water is the image, assuming that it is directly over the object? How deep would it appear to be if viewed from a point directly over the object?

6. Using the crown and flint glasses for which the data are given in § 204, find the angle of a flint prism which will be necessary to produce achromatization for the C and F lines for a 10° crown glass prism. Compute the deviations of the B, C, D, F, and G lines.

7. A crown glass prism of 15° is combined with a flint glass prism so that the D line is undeviated. What must the angle of the flint prism be and what are the deviations of the B, C, D, F, and G lines?

8. Draw a figure showing (*a*) each angle and distance which must be computed, (*b*) what formulas to use, and (*c*) what indices of refraction to use, in order to find the lateral displacement of the B and G rays from the incident ray of white light striking a plate of crown glass at an angle of incidence of 60°. The plate is 2 inches thick.

209. Refraction through Lenses: Types of Lenses. — There are two general *types of lenses, converging* and *diverging.* The former are thicker at the center than at the edges and produce therefore a greater retardation of the portion of a wave front passing through the center than of the portion passing through near the edges. With a converging lens the curvature of the wave front is diminished and may be reversed; a plane wave will be rendered converging, *i.e.* all portions will converge toward a common point or focus. Diverging lenses on the other hand are thinner at the center than at the edges and produce less retardation of the central portion of an advancing wave front. A plane wave will be rendered diverging.

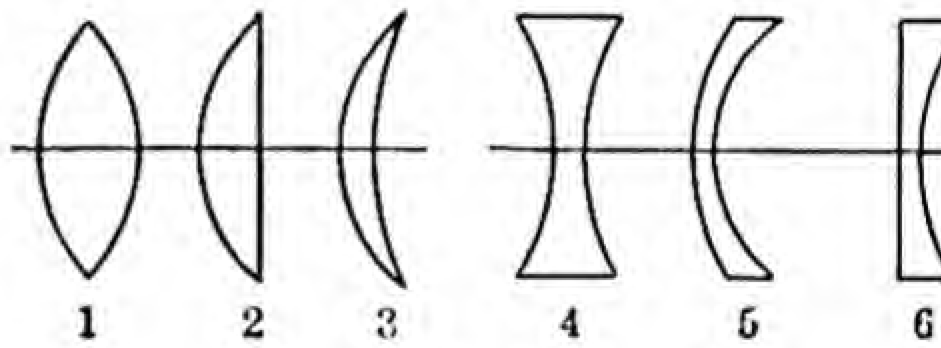


FIG. 184

Fig. 184 shows various types of lenses. 1, 2, and 3 are converging lenses; 4, 5, and 6 are diverging lenses. 1 is double convex, 2 is plano-convex, 3 is concavo-convex, 4 is double-concave, 5 is convexo-concave, 6 is plano-concave.

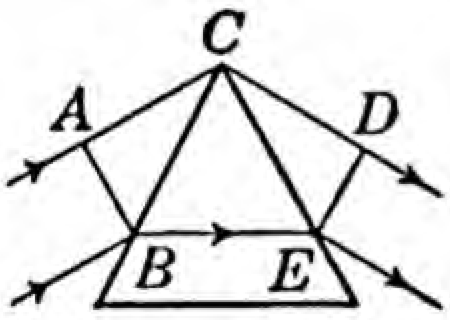


FIG. 185

210. Equivalent Air Path or Reduced Optical Path. — In Figs. 168 and 169 it was necessary to pick out two distances which two wave fronts, respectively, traveled in equal times. Since the paths were all in air, where the waves traveled with equal velocities, then the distances traversed were the same. When one of the wave fronts travels in a medium of refractive index n while the other travels in air, the distances covered in equal times are not

the same. In Fig. 185, consider the wave front AB which advances through a glass prism to the position DE. The portion at A travels the distance ACD, entirely in air, while the portion at B covers the distance BE during the same time. Let t be the time required for each of these motions, v be the velocity of the waves in air, v_g be the velocity in the glass, and $v/v_g = n$ the index of refraction of the glass. Then

$$vt = AC + CD,\ v_gt = BE,$$

and

$$AC + CD = \frac{v}{v_g} BE = nBE. \tag{313}$$

Hence if BE is the distance covered by a wave in a refractive medium in a given time, the *equivalent air path* or *the reduced air path* is nBE.

211. Conjugate Focal Points for a Converging Lens. — In Fig. 186 is shown a double convex lens. The surface FBG is a sphere whose center is at C_1 and whose radius is $C_1B = r_1$. The surface FEG is a sphere whose center is at C_2 and whose radius is $C_2E = r_2$. The lens forms a real image at I of the object at O. Two important wave fronts are shown, ABC and HEK, tangent to the lens.

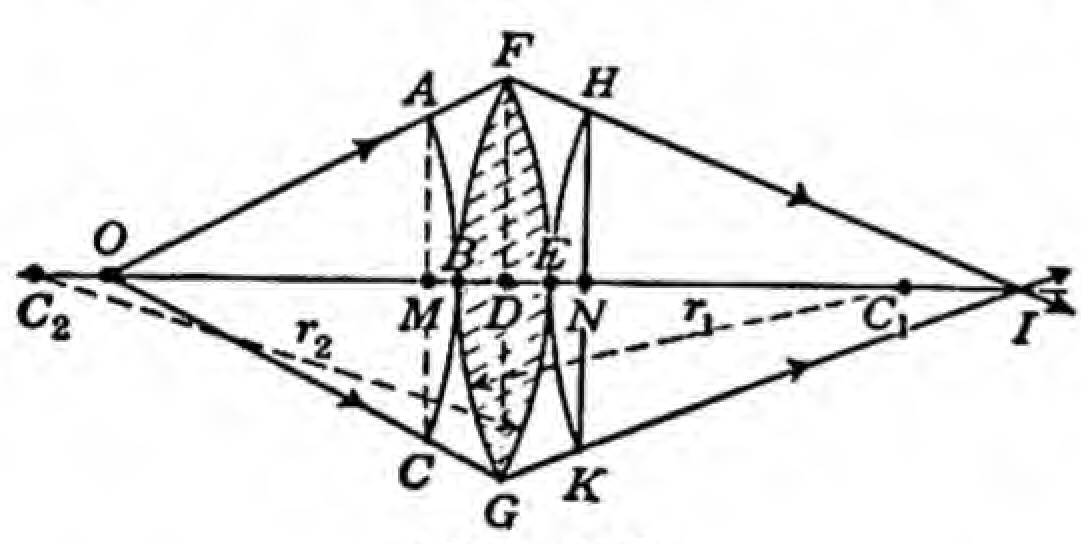


FIG. 186

While the wave at A advances to F and H in air, the portion at B advances to E through the glass. By Eq. 313,

$$AF + FH = nBE. \tag{314}$$

When the distances of the object and image are large compared to the size of the lens, then AF is nearly equal to MD and FH to DN. Under the same conditions, AM, FD, and HN are nearly equal. Eq. 314 must now be transformed into distances which represent sagittae of the arcs ABC, FBG, FEG, and HEK.

$$MB + BD + DE + EN = n(BD + DE),$$

and

$$MB + EN = (n - 1)(BD + DE).$$

Now expressing these sagittae in terms of the chords and radii and canceling, we obtain

$$\frac{1}{p} + \frac{1}{q} = (n - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right), \tag{315}$$

where p and q represent the distances of the object and image, respectively, from the lens. Just as in the case of the spherical mirrors, we define the distance from the lens to the image point of a very distant object as the principal focal length, f. Hence for $p = \infty$, $q = f$, and we obtain,

$$\frac{1}{q} = \frac{1}{f} = (n - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right).$$

We may write Eq. 315 as

$$\left.\begin{aligned} \frac{1}{p} + \frac{1}{q} &= \frac{1}{f}, \\ \text{in which} \qquad f &= \frac{1}{n-1}\frac{r_1 r_2}{r_1 + r_2}. \end{aligned}\right\} \tag{316}$$

212. Conjugate Focal Points for a Diverging Lens.— A spherical wave from O (Fig. 187) spreads out and the portion ABC falls upon a diverging lens. The diverging emergent wave front FHG has a curvature as if it had arisen from the point I, the virtual image position. While the portion of the wave at A has traveled through the lens to F, the portion at B has traveled to H, partly through air and partly through the lens. Applying the principle of Eq. 313 to this case, we get

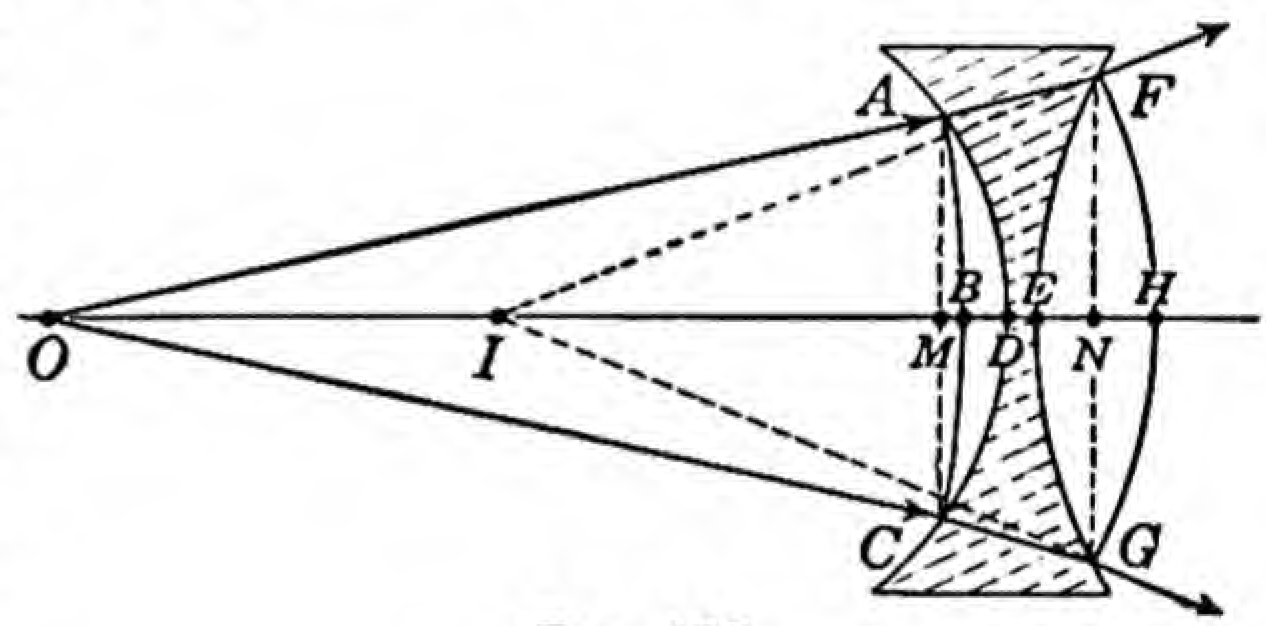


Fig. 187

$$nAF = BD + n\,DE + EN + NH, \tag{317}$$

where the angles subtended at O and I are very small, AM is nearly equal to FN, and AF is nearly equal to MN. So

$$n(MD + DE + EN) = BD + n\,DE + EN + NH.$$

This equation will now be changed so as to involve only distances which are sagittae of the arcs ABC of radius p, ADC of radius r_1, FEG of radius r_2, and FHG of radius q. First, the terms $n\,DE$ cancel each other.

$$n(MD + EN) = MD - MB + EN + NH,$$

and

$$(n - 1)(MD + EN) = NH - MB.$$

Expressing these sagittae in terms of the chords and the radii and canceling, we obtain

$$\frac{1}{p} - \frac{1}{q} = -(n-1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right), \tag{318}$$

which is identical to Eq. 315 in form but not in signs.

213. The Rule of Signs for Lenses. The Diopter. — It is obvious that Eq. 316 may be used for both converging and diverging lenses if some rule is adopted to make the proper terms negative. The rule is: *For the standard case, when the incident rays diverge and the transmitted rays converge, p and q are positive. For the opposite cases, p and q are negative, i.e.* when the incident rays converge, the object is virtual and p is negative; when the transmitted rays diverge, the image is virtual and q is negative. *The positive sign is used for r_1 or r_2 when the corresponding surface of the lens is such that the rays which enter are made more convergent upon passing through the surface. Otherwise r_1 or r_2 is negative.* An equivalent statement is: r_1 or r_2 is positive if the corresponding surface is convex when viewed from the medium of lesser refractive index and is negative if the surface is concave as viewed from the medium of lesser refractive index.

In Fig. 184 the signs for the radii of the left-hand surfaces, reading from left to right, are $+, +, +, -, +, \pm$. Likewise, the signs for the radii of the right-hand surface are respectively, $+, \pm, -, -, -, -$. If the algebraic sum of $(1/r_1 + 1/r_2)$ is positive, the lens is convergent; if negative, divergent.

In practical optometry a unit is needed which will indicate the amount of aid which a spectacle lens must give to the human eye. When the crystalline lens or the cornea of the eye is just slightly out of shape, only a "weak" lens is needed, — *i.e.* the center is very slightly different in thickness from that at the edges. For more defective eyes "stronger" lenses are required. The reciprocal of the focal length of a lens is a measure of the strength of the lens. The reciprocal meter, called the *diopter*, is taken as the unit of strength. Thus the "strength" or "power" of a converging lens whose focal length is 4 meters is said to be $+ 0.25$ diopters. A diverging lens of the same focal length has a power of $- 0.25$ diopters.

214. The Aberrations of a Single Spherical Lens. — When a lens is made of one single piece of glass with two spherical sur-

faces, there are several separate causes for poor image formation. These defects are not due to imperfections in the glass or in the grinding of the surfaces but are inherently geometrical properties of the sphere. The most important defects are spherical aberration, chromatic aberration, coma, astigmatism, curvature of field, and distortion.

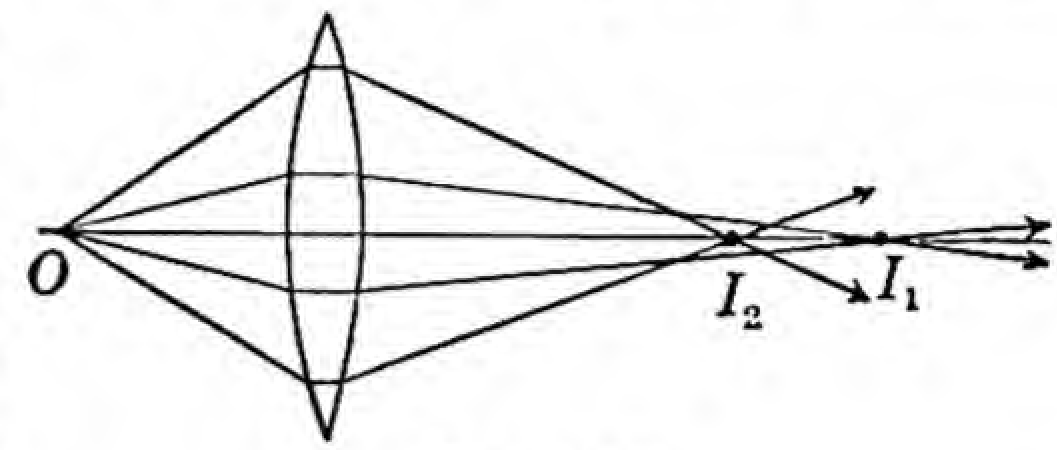


Fig. 188

(1) *Spherical Aberration.* — If the rays from a point source at O passing through the central portion of the lens form an image at I_1 (Fig. 188), then it is found that the rays passing through the outer zones of the lens will form an image at a point closer to the lens, such as at I_2. The intermediate zones have focal lengths covering separately all the intermediate space between I_1 and I_2. A screen placed at any particular place in the region from I_1 to I_2 will receive various sized annular images of O formed by the various zones, the total effect being a circular image whose size and whose distribution of illumination over the circle vary with the position of the screen. At no place, however, is a point image obtained.

(2) *Chromatic Aberration.* — Consider again a point source at O (Fig. 189), and in order to make negligible any effects due to spherical aberration, let a screen S cover up all but one narrow annular zone of the lens. Then, due to the dispersion of the lens substance, the violet rays from O will be bent more than the red rays so the image of O formed by the violet light will be formed at some point I_v nearer to the lens than the point I_r where the image is formed by the red rays. At I_v the image formed on a screen would be a circle, blue at the center and red at the edges. At I_r the circular image would be larger, the order of the spectral colors being reversed. At some intermediate point the center would be green or yellow with the outer edges a mixture of red and violet.

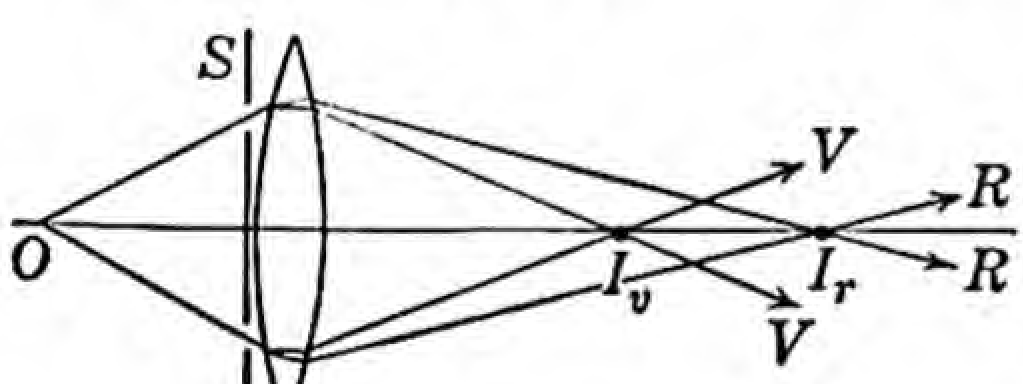


Fig. 189

(3) *Coma.* — Both coma and astigmatism show up in the image formation of a point which is not on the axis of the lens. Due to coma, the image consists of a sharp bright point with a comet-like

tail directed radially away from the axis of the lens. This defect becomes especially objectional when the distance between two images on a photographic plate is to be measured. Due to the lopsided nature of the image, the exact center cannot be determined.

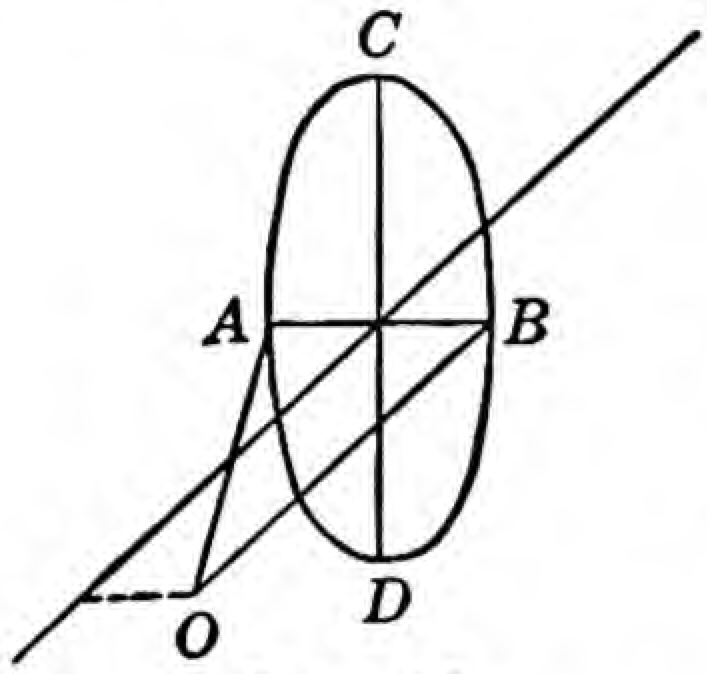


FIG. 190

(4) *Astigmatism.*—It is found that when the object is at O (Fig. 190), not on the axis, the rays such as OA and OB, in the plane containing the point O and the axis, cross, after passing through the lens, at a point which is farther from the lens than the point where rays such as OC and OD, in a plane at right angles to the plane OAB, meet. The total effect of this variation in focal length is shown in Fig. 191. For convenience in drawing, just the rays passing through a square portion of the lens are shown. The rays OA and OB after refraction cross at I_1. Further, OE and OF cross at I_3 and OG and OH cross at I_4. Thus, oblique rays from the point O form a line image I_3I_4. But projecting these rays further, they are seen to cross again in a line I_5I_6. Among these lines, may be

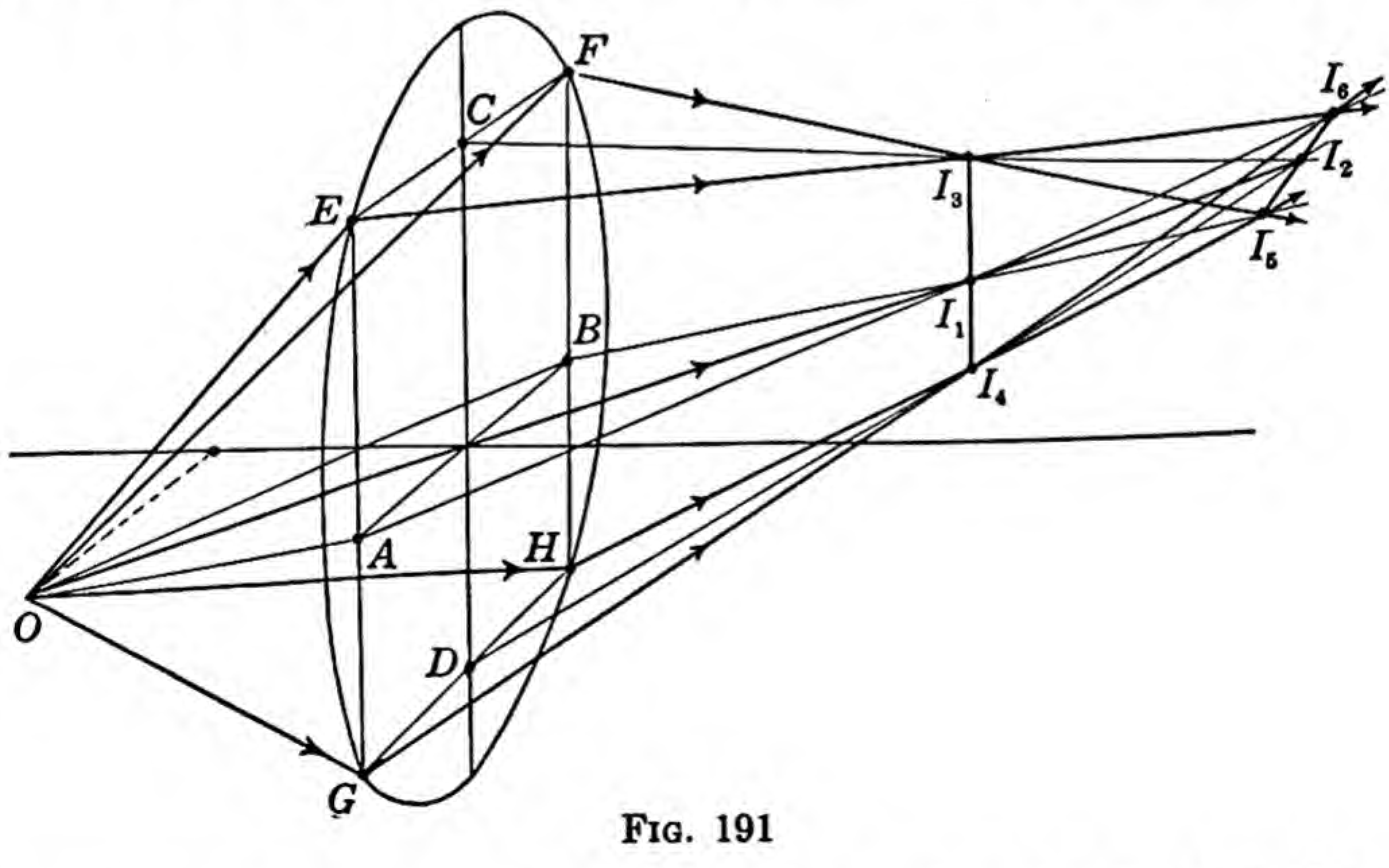


FIG. 191

seen the refracted rays of OC and OD crossing at a point I_2, further from the lens than I_1, as was stated to be true in Fig. 190. Astigmatism thus produces not a single image of a point object but two lines at right angles to each other. Using the whole circular

aperture of the lens, there is a place part way between the two line images where the image is circular. This is called the circle of least confusion and is the nearest approach to an image of the point O.

(5) *Curvature of Field.* — When an image of a flat extended object is formed, the image is not on a flat plane on the opposite side of the lens but is on a somewhat egg-shaped figure with its edges nearer the lens than its central portion.

(6) *Distortion.* — A single lens with spherical surfaces produces magnification which varies with the distance from the axis. So the image of an evenly divided scale would appear as a scale the size of whose divisions continually increases or decreases as the distance of the division from the axis increases.

All of the above defects exist simultaneously and account for the poor image formation of a single lens. Lens manufacturers must combine lenses of different kinds of glass and of different curvatures in such a manner that all these errors are reduced to a minimum or reduced to an extent which is negligible for the particular use for which the lens is intended. Lens systems used for visual work must be corrected for the middle region of the spectrum where the eye is most sensitive. Since the ordinary photographic plate or film is more sensitive in the green and blue, photographic lenses are corrected so as to give the least aberrations in the shorter wave-lengths. A photographic lens which is to be used for three-color photography, — red, green, and blue, — must be more highly corrected and is, therefore, a more expensive lens.

215. Combination of Lenses. — If the distance between two lenses is small compared with their focal lengths, the effective focal length of the combination is a simple combination of the two separate focal lengths. Let an object be distant p from a two-lens combination, the first lens having focal length f_1 and the second, f_2. For the first lens we have

$$\frac{1}{p} + \frac{1}{q'} = \frac{1}{f_1}.$$

The rays emerging from the first lens are incident upon the second lens. The distance q', with sign changed, is the object distance for the second lens when the distance between the lenses is small compared to q', because if q' is positive, the rays are convergent and

make a virtual object for the second lens, — and conversely if q' is negative. Then

$$-\frac{1}{q'}+\frac{1}{q}=\frac{1}{f_2}.$$

Adding these two equations,

$$\frac{1}{p}+\frac{1}{q}=\frac{1}{f_1}+\frac{1}{f_2}=\frac{1}{f}. \tag{319}$$

So the focal length, f, of a two-lens combination is seen to be

$$f=\frac{f_1 f_2}{f_1+f_2}. \tag{320}$$

The sign of f is determined algebraically by the sign of the right-hand member of Eq. 320 when the proper signs are used for f_1 and f_2.

216. Achromatic Combination. — In a similar fashion to the building of an achromatic prism combination (§ 208), we may find how to design a pair of thin lenses which will have the same focal length for two given wave-lengths for rays close to the axis. In Fig. 192 is shown such a combination. Let the surfaces from left to right have radii r_1', r_2', r_1'', and r_2''. As shown in the figure, $r_2' = r_1''$, but this need not be. A short air space may exist between the two component lenses provided the distance is small compared with the object and image distances. Let the single primed letters refer to the left-hand lens and the double primed letters to the right-hand lens. Applying Eqs. 319 and 320, we have the following condition that the combination of lenses have the same focal length for the C and F lines:

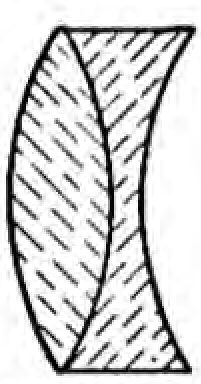


FIG. 192

$$\frac{1}{f_C}=\frac{1}{f_C'}+\frac{1}{f_C''}=(n_C'-1)\left(\frac{1}{r_1'}+\frac{1}{r_2'}\right)+(n_C''-1)\left(\frac{1}{r_1''}+\frac{1}{r_2''}\right)$$

$$=\frac{1}{f_F}=\frac{1}{f_F'}+\frac{1}{f_F''}=(n_F'-1)\left(\frac{1}{r_1'}+\frac{1}{r_2'}\right)+(n_F''-1)\left(\frac{1}{r_1''}+\frac{1}{r_2''}\right),$$

or
$$(n_C'-n_F')\left(\frac{1}{r_1'}+\frac{1}{r_2'}\right)=(n_F''-n_C'')\left(\frac{1}{r_1''}+\frac{1}{r_2''}\right). \tag{321}$$

The signs of the various radii are determined as in § 213. This equation holds true only as an approximation, the limitations being the same as those under which the simple lens formula holds.

217. Linear and Angular Magnification. A Single Lens as a Magnifier. — If an object is placed at E (Fig. 193), just inside of the principal focal point F of a converging lens, an enlarged virtual image will be produced at E' some distance beyond F. It is seen that the rays AT and CD, originating from the point A, are divergent. If the object is moved slightly toward F, only the ray AT changes and just a slight motion of E toward F will throw the image position very much farther away from the lens. The eye is placed close to the right-hand side of the lens and receives the light diverging from it. It is found that the eye can see the image clearly for all of its positions from 25 cm. to infinity. When the image is at infinity, the normal eye is in a completely relaxed condition. For nearer positions, up to 25 cm., the eye can accommodate itself without undue strain on the muscles of the eye. For the average person of 30 to 40 years of age, objects can be seen most clearly with the greatest ease if they are approximately 10 inches or 25 cm. from the eye. *This distance of 25 cm. is called the standard distance of distinct vision.* It is rather arbitrary since younger people have a smaller and older people a larger distance of distinct vision than the standard.

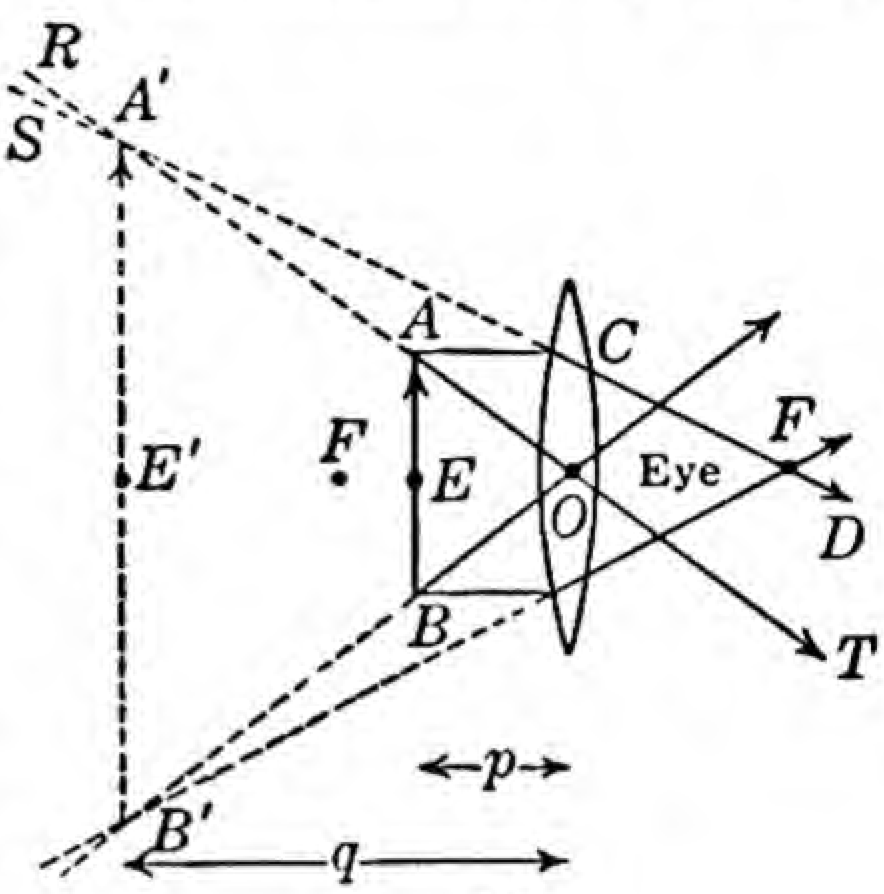


FIG. 193

Examination of the ray construction for converging or diverging lenses with real or virtual images shows that the ratio of the size of the image to the size of the object is the same as the ratio of the image distance to the object distance from the lens. *When the images are real, we define the linear magnification of a lens or system of lenses as the ratio of the final image size to the object size.* As will be shown in the next paragraph, such a definition for virtual images, which are viewed by the eye, is not definite and of no practical value.

As the image $A'B'$ recedes to infinity, its size increases proportionately. Its linear magnification becomes enormously large and yet, as we shall see shortly, its size as judged by the size of the retinal image changes but slightly. The size of the retinal image in the eye is directly proportional to the angle subtended by the rays entering the pupil. Our entire conception of magnification depends

upon the ratio of the size of the retinal image formed with the aid of the optical instrument to its size when viewed by the unaided eye. It is thus obvious that *for virtual images* we must speak of angular magnification. *We define angular magnification of a visual instrument as the ratio of the angle subtended by the rays emerging from the instrument to the angle subtended at the eye by the rays coming directly from the object.* When the object is movable, this latter angle is taken when the object is at the standard distance of distinct vision. When the distance between the object and observer is fixed, the angle is computed as the ratio of the height of the object to the distance between the object and observer.

We may examine how the angle $A'OB'$ changes as the image recedes from 25 cm. to infinity. As the object is moved from E toward F, the path ACD remains unchanged, and, therefore, the virtual ray SC is unchanged. But the ray AOT is changed so that the virtual ray RA intersects SC at a point more distant from the lens. Thus the image of point A moves along the ray SC and not along RA. Had it moved along RA, its angular dimension subtended at the lens would be unchanged, but its actual path along SC makes the angle subtended increase as the image moves in toward the lens. Then it is plain that the largest magnification will be obtained when the image is nearest the lens. This is obtained by moving the object nearer to the lens. The practical limit to this increase is reached when the image reaches the distance of distinct vision.

The angular magnification when the image is 25 cm. from the eye and the lens will now be computed. The angle subtended by the object with the unaided eye must be $AB/25$. The angle subtended by rays emerging from the lens is

$$\frac{A'B'}{E'O} = \frac{A'B'}{25}.$$

From similar triangles, we obtain

$$\frac{A'B'}{25} = \frac{AB}{p}.$$

Hence the angular magnification (A.M.) is

$$\frac{\dfrac{A'B'}{25}}{\dfrac{AB}{25}} = \frac{A'B'}{AB} = \frac{25}{p}. \tag{322}$$

But by Eq. 316 $\quad \dfrac{1}{p} - \dfrac{1}{25} = \dfrac{1}{f},$

so $$\text{A.M.} = 25\left(\frac{1}{f} + \frac{1}{25}\right) = 1 + \frac{25}{f}. \tag{323}$$

As AB approaches F and the image recedes to infinity, p approaches f and Eq. 322 gives A.M. = $25/f$, which is slightly less than the maximum value. If the eye is not held very close to the lens, then the angles subtended at the eye are less than those computed above and therefore the magnification will be less.

A further advantage is to be had by placing the eye close to a simple magnifier, — the portion of the object which may be viewed without distortion is much greater than when the eye is farther away.

218. The Compound Microscope. — A compound microscope usually consists of two lenses. The objective (lens) CD (Fig. 194) has a very short focus so that the object AB can be placed outside

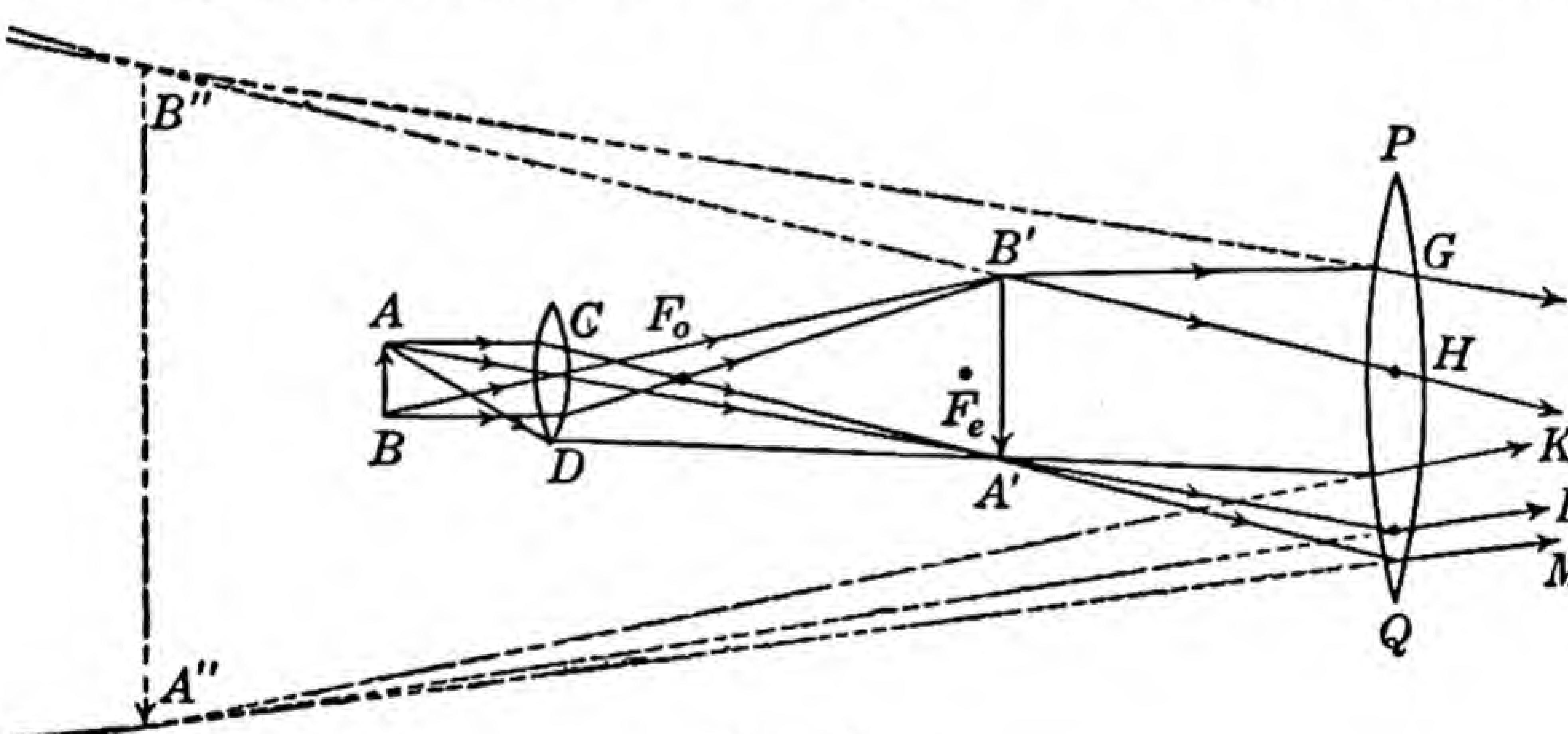


FIG. 194

its principal focus but still very near it. Then the real image $A'B'$ formed at a distance from the lens will have a very large linear magnification. The eye lens PQ must be placed so that the real image $A'B'$ is formed inside its principal focal point in order that an enlarged virtual image $A''B''$ be formed. In the lower half of the eye lens are shown some of the actual rays from A which pass through the objective, cross at A', and are refracted by the eyepiece along K, L, and M. The geometrical construction of the rays K, L, and M is not simple, but the image of $A'B'$ may be located by the following fictitious method. Consider the point B' as an object giving out light in all directions. Draw the ray

$B'G$ which is refracted down through the principal focus of the eye-piece. Then draw the ray $B'H$ passing through the center of the lens. These divergent rays when projected backward diverge from the image point, B''. The rays $B'G$ and $B'H$ do not exist. Only a narrow bundle of rays from C to D actually passes through B' similar to that drawn through A'. However, if the objective CD were large enough, then the rays $B'G$ and $B'H$ would exist and therefore they may be used for locating the image.

The magnification of the microscope is the product of the separate magnifications of the objective and eye-piece. Since $A'B'$ is a real image we may consider the linear magnification. Since in practice AC is very small compared to CA', CA' may be considered as an infinite distance and so AC must be nearly equal to the principal focal length of the objective f_o. Let L be the distance between the two lenses and f_e be the focal length of the eye-piece. Then the distance CA' is nearly $(L - f_e)$ and so the first magnification is $(L - f_e)/f_o$. If the image $A''B''$ is focussed 25 cm. from the eye, a further magnification of $1 + 25/f_e$ is obtained by Eq. 323. If $25/f_e$ is fairly large and again if the final image is formed at infinity, we have $25/f_e$ for the second magnification, and so for the total angular magnification we have

$$\text{A.M.} = \frac{(L - f_e)25}{f_o f_e}. \tag{324}$$

219. Astronomical and Terrestrial Telescopes. — A telescope differs from the microscope because of the restriction placed on the object position. In a microscope the object may be placed wherever convenient. A telescope is used to view distant objects. The objective must form as large a real image as possible and the eye-piece must form as large a virtual image of it as possible. Such a telescope is shown in Fig. 195. For simplicity, the eye-piece is shown as a single lens. There are many types of eye-pieces containing two or more separate lenses. Fig. 195 is lettered the same as Fig. 194. The two are optically identical but must be widely different in proportions. The distance

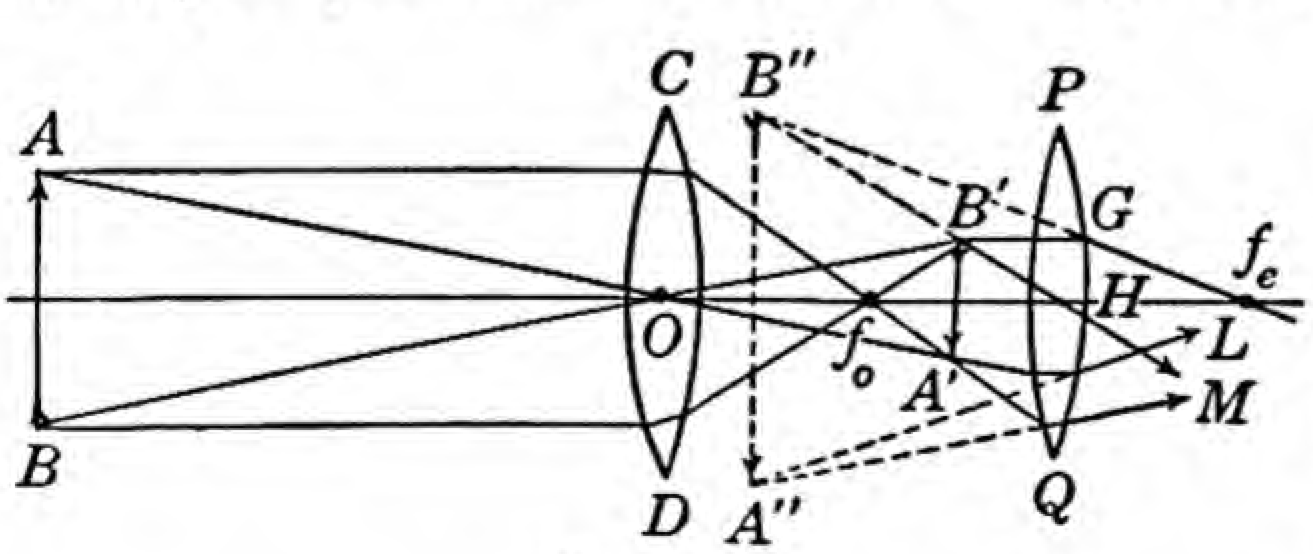


FIG. 195

AC (Fig. 195) is very large. Using an objective, CD, with as long focal length as convenient, there results a real inverted image $A'B'$ which is many times smaller than the object size, but which is as large as can be obtained. The eye-piece must then magnify that image as highly as possible. In some cases the eye-piece consists of a compound microscope the objective of which forms an enlarged real image of $A'B'$ and whose eye-piece further magnifies that image. Thus in order to get the great magnification the objective lens must have as long a focal length as possible and the eye-piece must have as short a focal length as possible.

The angular magnification is obviously the ratio of the angle subtended at the eye by the image $A''B''$ to the angle subtended by the object at the eye. Since both the objective lens and the eye are very far from the object, we may take the angle AOB as the angle subtended at the unaided eye by the object. Angle AOB = angle $A'OB'$ which in turn is nearly equal to $B'A'/f_o$. Angle $B''HA''$ is equal to angle $B'HA' = B'A'/f_e$ approximately. So

$$\text{A.M.} = \frac{f_o}{f_e}. \tag{325}$$

The telescope which has just been described produces a final image which is inverted with respect to the object. Telescopes of this type are used in astronomical work where the inversion of the image is not objectionable. Hence such instruments are called astronomical telescopes. A telescope designed to produce an erect image is called a terrestrial telescope. An astronomical telescope may be changed to a terrestrial telescope by replacing the single lens eye-piece by a somewhat longer two lens system. The first lens, such as PQ, is placed so that the image $A'B'$ is outside its focal length. Then a real inverted image of $A'B'$ is formed, which image is therefore erect. The second lens of the eye-piece is placed as a simple magnifier with respect to this erect image.

220. The Galileo Telescope. — In the telescope devised by Galileo, a diverging lens is used for the eye-piece. The lens is placed within the focal length of the objective and thus intercepts the converging rays before the image is formed by them. In Fig. 196 is shown the effect of such a lens on converging rays. In the first figure the lens is placed so that its principal focus is in the same plane in which the incident rays would cross. The emergent rays are parallel, for if the rays are reversed, then by definition the

parallel rays on emerging diverge from the point A' which is in the principal focal plane of the lens.

In the second figure the lens has been moved nearer the point A' so that its principal focus is beyond A'. In this case the curvature of the incident waves is so great that the lens cannot even make the wave fronts plane, so the rays cross at a point A more distant from the lens than A'.

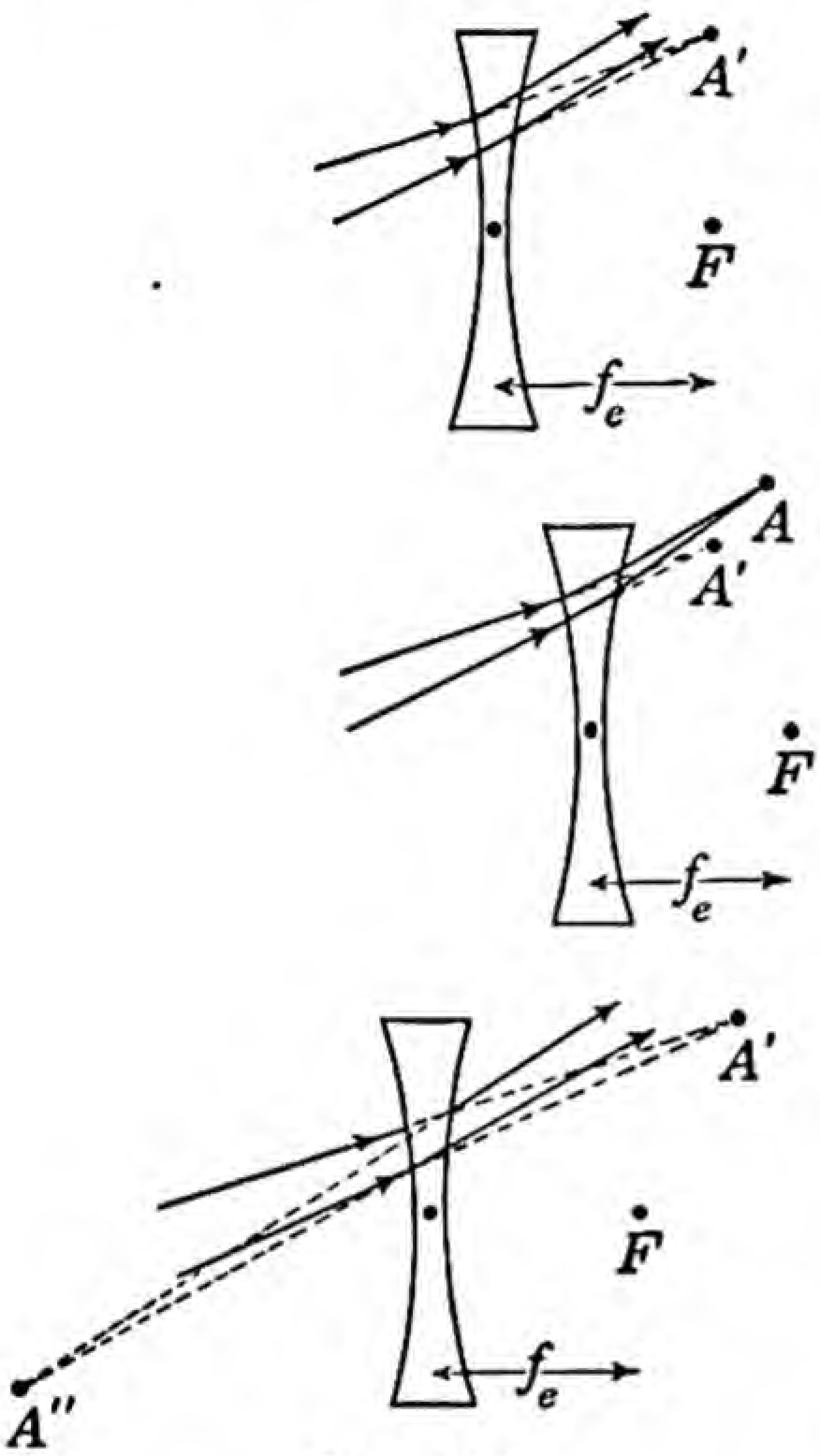


FIG. 196

In the third figure the lens is moved so that A' is beyond the principal focus. The incident wave front has a small enough curvature that the lens reduces it to zero and reverses it, so the emerging rays are divergent and form a virtual image at A''. This latter case is the arrangement for the eye-piece of the Galilean telescope.

Fig. 197 shows the complete arrangement of the telescope. For convenience the object is not shown. It is so far away that the rays from any point of it are considered parallel. The rays from the top of the object are labeled A and the rays from the bottom of the object, B, as in Figs. 195 and 196. The rays A after passing through the lens converge toward a point A' which is slightly beyond the principal focal point of the objective, F_o, and likewise the rays B converge toward B'. The eye lens intercepts these rays and $A'B'$ is a virtual object for the eye lens. The eye lens is placed so that its principal focal point, F_e, is just inside of the virtual object position. Then as shown in Fig. 196 the incident rays are made divergent so that an erect virtual image $A''B''$ is formed. As in the other diagrams, the A rays are carried completely through the system although no construction is shown to give the amount of deviation produced

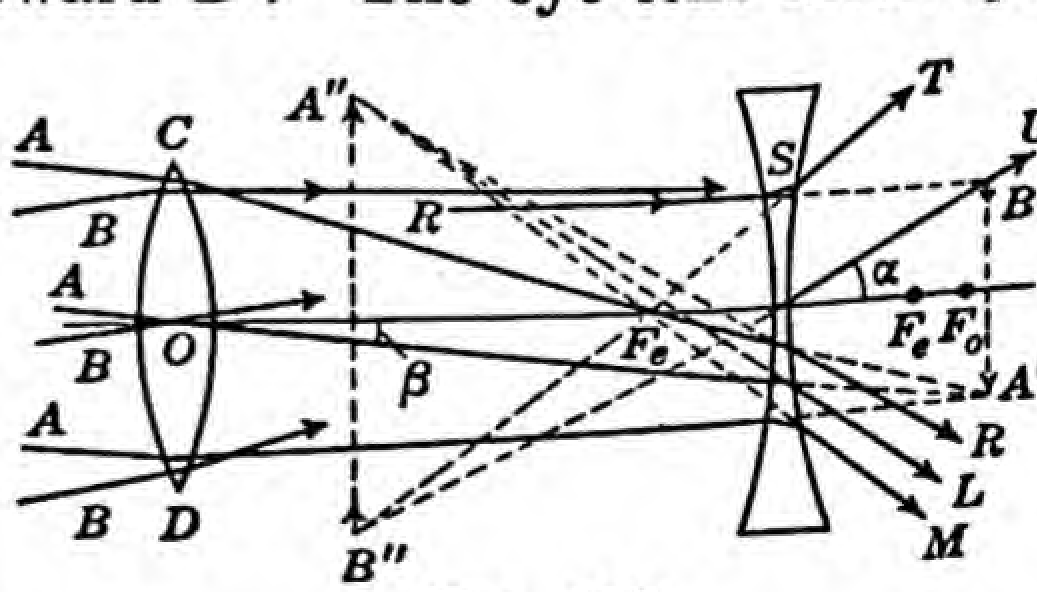


FIG. 197

by the eye-piece. However, the ray construction is shown which locates the image B'' of the virtual object B'. The ray RST is one of the rays which actually exist, but the ray $B''U$ is fictitious because no such ray could exist unless the objective lens were larger.

For all practical purposes we may consider the image $A'B'$ to be at the point F_e and if the eye-piece is moved until the image $A''B''$ is at infinity, then F_o lies practically in $A'B'$. The angular magnification is the ratio α/β,

$$\alpha = \frac{A'B'}{2f_e} \text{ and } \beta = \frac{A'B'}{2f_o},$$

so,

$$\text{A.M.} = \frac{f_o}{f_e}. \tag{326}$$

This type of telescope is used in most opera glasses because the system is short due to the position of the eye-piece. The image $A''B''$ is usually 25 cm. from the eye-piece so that if large magnification is desired, the distance CS is made very large and the image $A''B''$ will lie between the two lenses relatively close to S.

PROBLEMS

1. In Fig. 184, if the lenses 1, 4, and 5 were made of glass whose $n_D = 1.5000$, what would be the magnitude and sign of the focal lengths if the left-hand surface had a radius of 10 cm. and the right-hand 5 cm.? What is the power of each lens in diopters?

2. A far-sighted person can see without eye strain things one meter away from his eyes. What kind of lens will enable him to see things at 25 cm. from his eyes? What must be the power of the spectacle lenses in diopters if they are placed one centimeter from his eyes?

3. A near-sighted person who can see things clearly when placed 7 cm. from the right eye and 9 cm. from the left eye, is to be fitted with spectacles so that he may see easily things placed 25 cm. from his eyes. Find the powers of the two lenses which are to be placed one centimeter from his eyes.

4. If the first lens in Problem 1 is placed under water, what does its focal length for the D line become?

5. A flint-glass lens whose surfaces have radii of + 10 cm. and + 20 cm. is combined with a crown-glass lens whose surfaces have radii of + 5 cm. and − 10 cm. What is their combined focal length for the Fraunhofer C line?

6. An object 1 in. high is placed 10 in. in front of a converging lens of 5 in. focal length. Beyond the lens is placed another similar lens. Where is the final image, is it real or virtual, and how large is it, when the distance in inches between the lenses is, in turn, 25, 15, 12, 8, and 0?

7. In Problem 6, if the second lens has a focal length of − 5 in., locate the image for the following lens separations: 15, 7, 5, 3, and 0 inches.

8. If a source and a screen are farther apart than four times the focal length of a certain converging lens, show that there will be two positions of the lens at which a real image of the source will be formed on the screen. Prove that if l is the distance from source to screen and a the distance between the two positions of the lens, the focal length of the lens is given by $(l^2 - a^2)/4\,l$. (Draw a diagram and make use of the symmetry of the figure.)

9. Explain why objects look smaller when inside of an empty thick-walled bottle and yet look larger when the bottle is filled with water.

10. For diamond, $n_D = 2.4173$. What is the critical angle? What bearing does this have on the brilliance of a cut diamond?

11. A ray of monochromatic light is incident internally on a face of a glass prism so that it is just totally reflected from a portion of it which is moistened with a liquid. The angle of incidence is 60° and the index of refraction of the prism for light of the same wave-length is 1.5624. What is the index of refraction of the liquid for that wave-length?

12. An achromatic lens is to be made having the same focal lengths for the D and G lines. The converging lens has equal radii of 30 cm. The diverging lens is ground to be cemented in contact with the converging lens. What must be the radius of curvature of its other face? Take the data in § 204.

Interference

221. General Statement. Young's Experiment. — Consider two sources, S_1 and S_2 (Fig. 198), which are emitting a series of waves of the same wave-length and in phase with each other. The full lines represent the positions of the crests of the waves at a given instant and the dotted lines the corresponding troughs. At the points indicated by the circles, crests coincide with crests or troughs with troughs so that along the line of circles the amplitude of vibration is double that due to either set of waves taken separately. Along the lines indicated by the succession of squares, crests meet troughs and troughs meet crests and the medium is not displaced. Thomas Young, in 1802, called this phenomenon interference. The word interference now includes all phenomena dealing with the combined effect of two or more sepa-

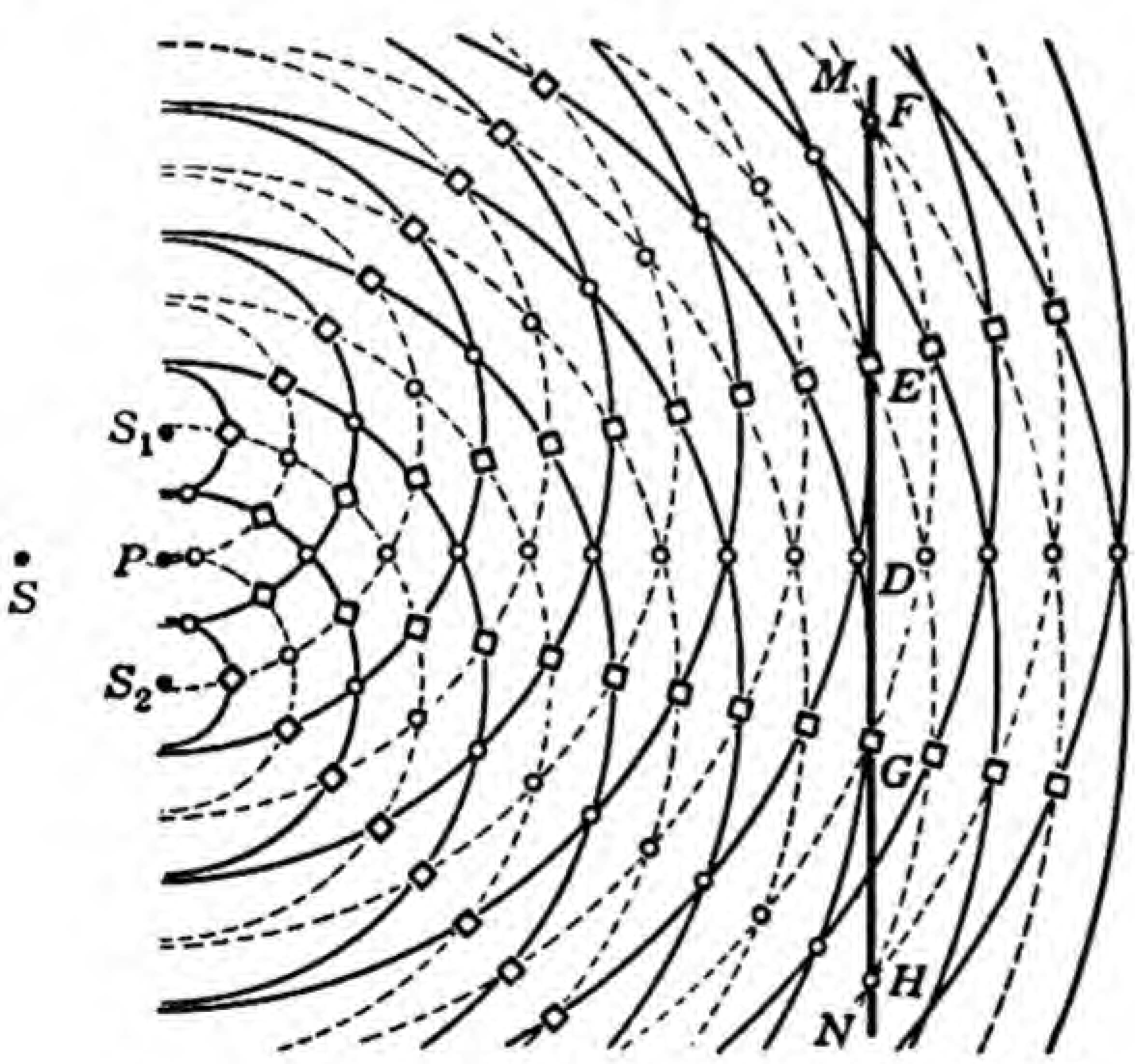


FIG. 198

rate sets of waves. Young described an experiment for showing the above case for light waves. In order to get two sources which are always in phase, he placed a light source at S and at S_1 and S_2 had two narrow slits in an opaque screen. When a screen MN was placed at some distance from the slits, alternate bright and dark streaks were seen on it corresponding to the points indicated by the circles and squares respectively. Young used this experiment to prove that light consists of a wave motion.

By counting the number of waves between the sources S_1 and S_2 and the points D, E, F, G, H, etc., we find that the distances S_1D and S_2D consist of the same number of wave-lengths; S_1E and S_2E differ in length by one half wave-length, as do also S_1G and S_2G; the distances S_1F and S_2F (also S_1H and S_2H) differ by one whole wave-length. In general, we find that for a point where the light intensity is a maximum, the difference in light paths from the two sources is $n\lambda$ where n is an integer. Likewise, for any point of zero intensity, the path difference is $(2n - 1)\lambda/2$, *i.e.* an odd multiple of a half wave-length.

Let P be the midpoint between S_1 and S_2. The student is expected to show for the n^{th} bright line on either side of the central line at D that $n\lambda = a \sin \theta$ where a is the distance between the two slits and θ is the angle between the line PD and the line from P to the n^{th} bright line.

222. The Bi-prism, Bi-mirror, and Bi-lens Methods of Obtaining Interference. — Fresnel placed a source at S behind a bi-prism P (Fig. 199), whose two slanting faces made a very small angle with each other. Light rays from S striking the upper half of the prism are refracted into the region between the lines AG and CH so that they appear to come from the point S_1.

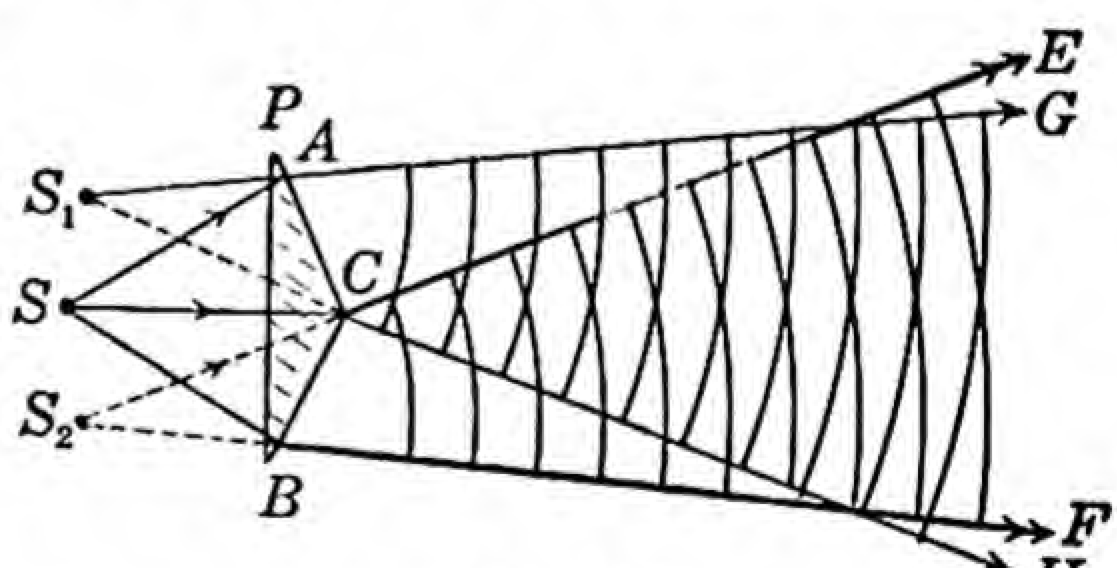


FIG. 199

The rays striking the lower half of the prism are refracted into the region between the lines CE and BF so that they appear to come from the point S_2. Since these rays originate from a single source, they will be in phase and interference patterns will be obtained in the region where the light from the two virtual sources overlaps just as explained in § 221.

Fresnel also produced interference by reflecting light from two plane mirrors whose planes were slightly inclined to each other.

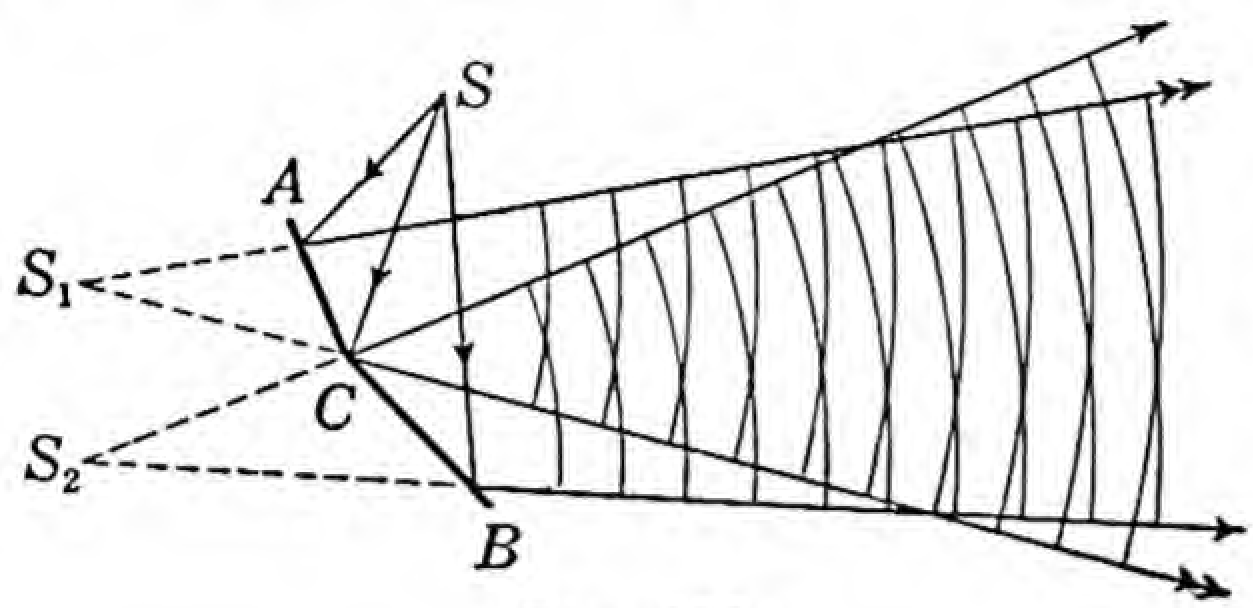


FIG. 200

A source at S (Fig. 200) is imaged in the two mirrors at S_1 and S_2 respectively. Interference bands will be produced in the overlapping region of the reflected beam.

Billet produced two virtual sources as shown in Fig. 201. A lens is split in two and the two halves slightly separated. Each half of the lens forms its own image of the source. The rays after crossing at S_1 and S_2 diverge and in the overlapping region interference bands will be produced.

The student may show that in all four methods described above, the two sources S_1 and S_2 must be very close together in order that the interference bands due to visible light be far enough apart that their separations be measurable.

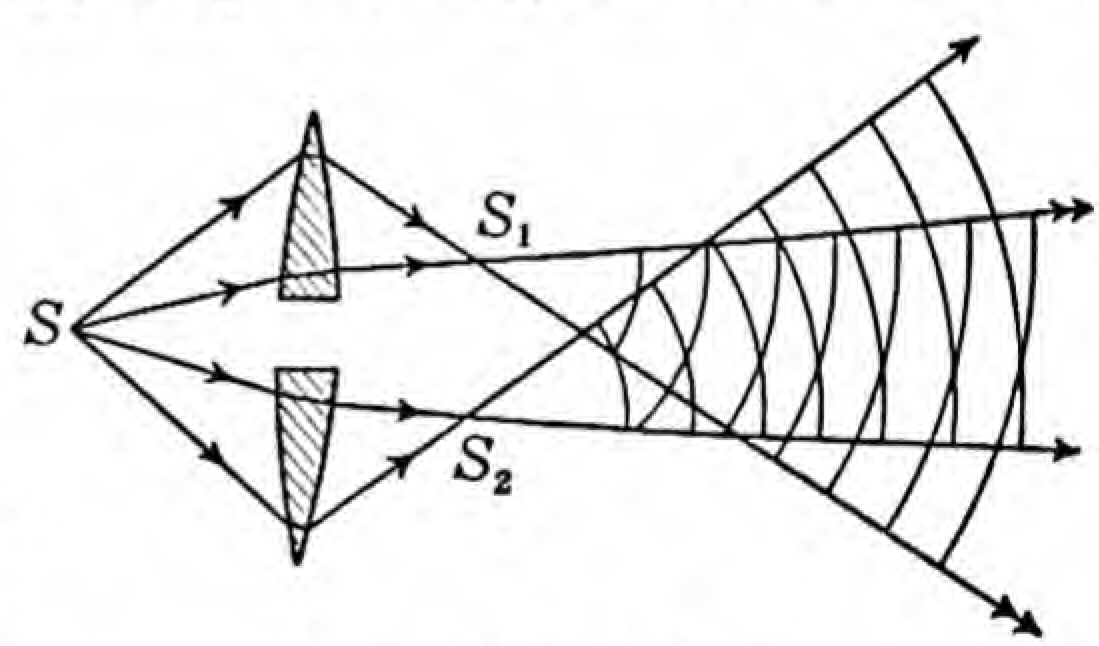


FIG. 201

223. Newton's Rings.—Newton discovered that if a spherical lens with a very small curvature was placed on a plane glass surface, then when light was reflected from or passed through them, alternate dark and bright circles surrounded the point of contact. The circles may be seen no matter at what angle the light falls on the system, but for purposes of measurement the rays should strike the surfaces normally. Let a monochromatic source S (Fig. 202) be placed at the focus of the lens P so that the rays emerge parallel. Let these rays strike a thin sheet of glass AF at a 45° angle of incidence. Most of the light will pass through the glass, but a certain amount will be reflected down to the lens L. Some of the light will be reflected from the lower surface while some will pass on through and be reflected from the upper surface of the plane plate of glass. Consider the ray FG. Since the curved surface of the lens is almost flat, the ray reflected at G will be almost along GF. The remainder of the light will pass on to H where a portion is

reflected back on itself. The two reflected rays, GF and HF, will be out of phase by an amount depending on their distance from B. Part of the light traveling along the ray HGF will pass through the glass at F and be seen by the observer above. In fact, wherever the two reflected rays are in phase, a whole ring of light, with its center at B, will be seen. At other points where the two reflected rays meet with opposite phases and annul each other, a dark ring will be produced. Successive concentric dark and bright rings result.

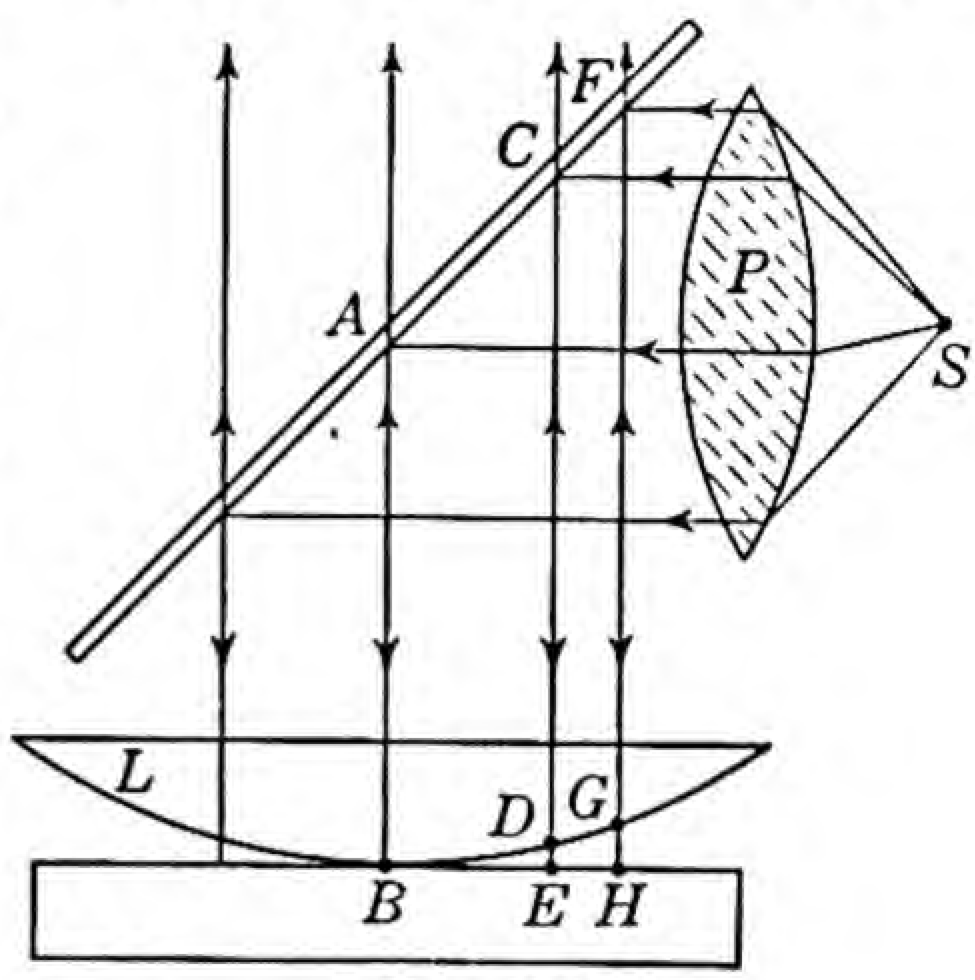


FIG. 202

It might seem at first thought that at B the two pieces of glass would be in contact so that the reflection at the lens surface would not be from glass to air. Nevertheless it is a fact that there are always many air molecules between the lens and the plate even at B. Special methods must be employed to remove the air even from two accurately ground optically plane glass surfaces in order to avoid reflections at the surface, which will occur if any air remains. It might then seem that the two reflected rays from B, having no difference in path lengths, would be in phase and the central spot would be bright. On the contrary, this central spot is found to be black. We must therefore conclude that the two sets of reflected waves are 180° out of phase. Now it will be remembered that at the end of a closed organ pipe a wave was found to be reflected with a 180° change in phase while at the end of an open organ pipe a wave was reflected with no change in phase. In a similar manner, the ray reflected at the lower surface of the lens (glass-to-air) suffers no change in phase while that reflecting from the upper surface of the plate (air-to-glass) undergoes a change in phase of 180°.

The wave-length of the light used may be experimentally determined as follows. The distance GH is seen to be equal to the sagitta of a chord of length 2 BH. Since the lens must be very flat in order to have the rings visibly separated, its radius of curvature must be very large and we may use Eq. 294. Remembering that reflection causes a difference in phase of $\lambda/2$ in the two re-

flected waves, we see that, in order to produce bright rings, there must be an additional phase difference of an odd multiple of a half wave-length. Hence for the first bright ring the path GHG $(= 2\,GH)$ must equal $\lambda/2$. For the sagittae of the various bright rings we may write,

$$2\,x_1 = \frac{\lambda}{2};\ 2\,x_2 = \frac{3\,\lambda}{2};\ 2\,x_3 = \frac{5\,\lambda}{2},$$

or in general,

$$2\,x = \frac{(2\,n - 1)\lambda}{2},$$

where n is the number of the bright rings counted from the center. Letting D be the measured diameter of the n^{th} ring,

$$x = \frac{D^2}{8\,R} = \frac{(2\,n - 1)\lambda}{4},$$

or

$$\lambda = \frac{D^2}{2(2\,n - 1)R}. \tag{327}$$

The radius of curvature of the lens must be measured by an auxiliary experiment.

224. Michelson's Interferometer. — Michelson devised an apparatus for obtaining interference fringes by which accurate measurement of wave-lengths may be made. The arrangement is shown in Fig. 203. The plates A and B are identical in thickness. They are usually cut from the same piece of glass whose two faces are ground optically flat and optically parallel, *i.e.* no spot on either surface is higher than any other by as much as one tenth of a wave-length of visible light and the distance between opposite faces is everywhere alike within one tenth of a wave-length. The mirrors C and D must have their front faces optically flat and are silvered on the front faces. The face of plate A next to plate B is silvered thinly so that about half of the light is reflected and the other half transmitted. Light from S is made parallel by a lens and falls on plate A, half of it being reflected toward plate B and half passing through to plate D. The reflected portion passes through B, is reflected from C, and then passes through B and A to the observer at O. The portion which passes through A is reflected from D,

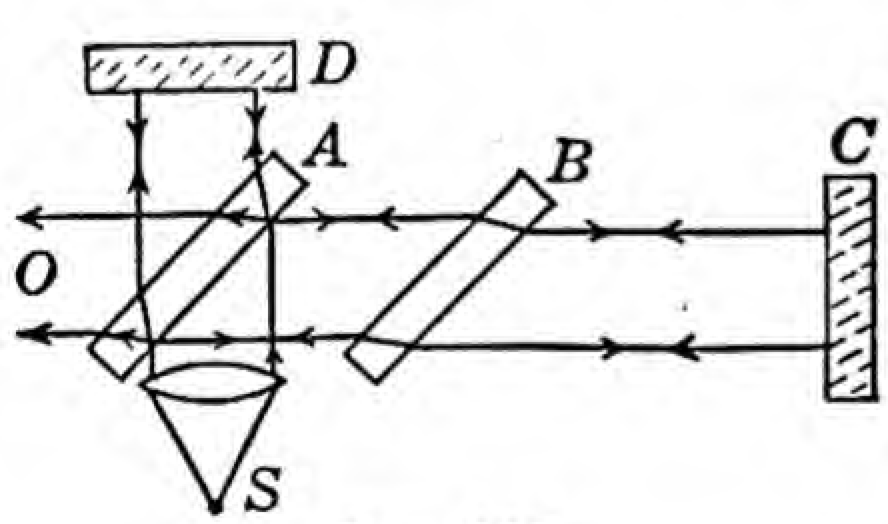


FIG. 203

passes through A to the half-silvered face, is reflected at that face, and passes out to O along the same path as the light coming from C. By the insertion of plate B into the system both paths are made alike, consisting of three passages through equal thicknesses of glass and two reflections. Although the reflection from C is from an air-glass surface and the reflection inside A is from a glass-silver surface, it has been shown experimentally that there is no phase difference introduced in the two beams by these two types of reflection. The observer at O sees a series of bright and dark bands (or fringes) which, according to the adjustment of the mirror, may be either a set of straight lines or a set of concentric circles. The mirror C is mounted on a very accurately ground screw so that it may be moved toward or away from B by accurately measured amounts. As C is moved, the fringes shift one way or another (in case of circular fringes they converge to the center or emerge from the center). When one fringe has moved from its position to that formerly occupied by the fringe next to it, the mirror C has moved exactly one half of a wave-length of the light used. By counting the passage of a large number of fringes and measuring the number of turns of the screw, exceedingly accurate measurements of wave-length may be made.

Using three lines of the cadmium spectrum, Michelson has determined the number of their wave-lengths required to make the length of the standard meter of Paris. There are reasons for believing that these measurements probably form a more permanent standard than the meter stick and certainly give a more indestructible one. The interferometer is used for many types of work where extremely small displacements are to be measured.

225. Energy Relations in Interfering Waves. — From a given steady source energy radiates at a definite rate and past any point in the space energy flows at a certain rate. When a second source is placed nearby, there may be, as we have seen, regions of zero energy and regions of increased energy. Since we know of no failure of the principle of Conservation of Energy, we conclude that in the process of interference, the energy of the two waves is not destroyed at the points of zero energy, but that the energy which would have appeared at those points has been transferred to the points where the amplitude of vibration is increased.

At those points where the two waves meet in phase the displacement is doubled. We shall now investigate how much the energy

of vibration is changed by a change of amplitude. We know that in the case of a mass moving with simple harmonic motion the energy of the mass is all potential when at its maximum displacement, and all kinetic when it is passing through its rest position. Those two energies must be equal in magnitude. Furthermore, at any intermediate position the sum of the kinetic and potential energies possessed by the particle is equal to either of the above-mentioned quantities of energy. The maximum potential energy may be computed by $W = \int F\,ds$. But $F = -cs$, so $W = \frac{1}{2}cs^2$ (§§ 18 and 48). Thus we see that the energy of a particle vibrating with Simple Harmonic Motion varies as the square of the amplitude, and hence in a place of maximum vibration in interference between two sources, the energy is four times as much as would exist due to either source separately.

Diffraction

226. General Conditions. — Under certain conditions it is observed that there is some illumination even in the geometrical shadow of an opaque object, and that there exists outside of the shadow a series of bright and dark fringes. It is also found when opaque screens are properly placed in a beam of light that, instead of causing the light intensity to decrease at certain points, the intensity may be increased many fold. The satisfactory explanation of these phenomena proved to be one of the triumphs of the wave hypothesis.

All diffraction phenomena, as will be seen in succeeding sections, occur when some obstacle is placed in front of a source of light, cutting off part of the beam. The study of the resultant effect produced by the remaining portion of the beam resolves itself into the study of the interference produced by the separate wavelets originating at each point of the remaining portion of the wave front. In fact, *we may call diffraction the effect produced by interference of the waves originating from portions of the same wave front.*

227. Half-Period Zones and the Zone Plate. — Consider a series of plane wave fronts such as MN, RS, and UV (Fig. 204). Let us investigate the effect of various parts of one of these wave fronts, UV, in producing illumination at the point P. Let d be the distance of P from the wave front. With P as a center describe

spheres of radii $d + \lambda/2$, $d + \lambda$, $d + 3\,\lambda/2$, etc., which cut out on the wave front circular zones of radii OA, OB, OC, etc. It is left as an exercise for the student to prove that the areas of the zones thus made increase only very slightly with increasing distance from P. Considering every point on UV to be a new source of light, according to the method of Huyghens, we see that the waves from A will reach P just $\lambda/2$ out of phase with the waves from O. Waves from all other parts of this first zone will reach P with phases differing from that of O by amounts varying from $\lambda/2$ to

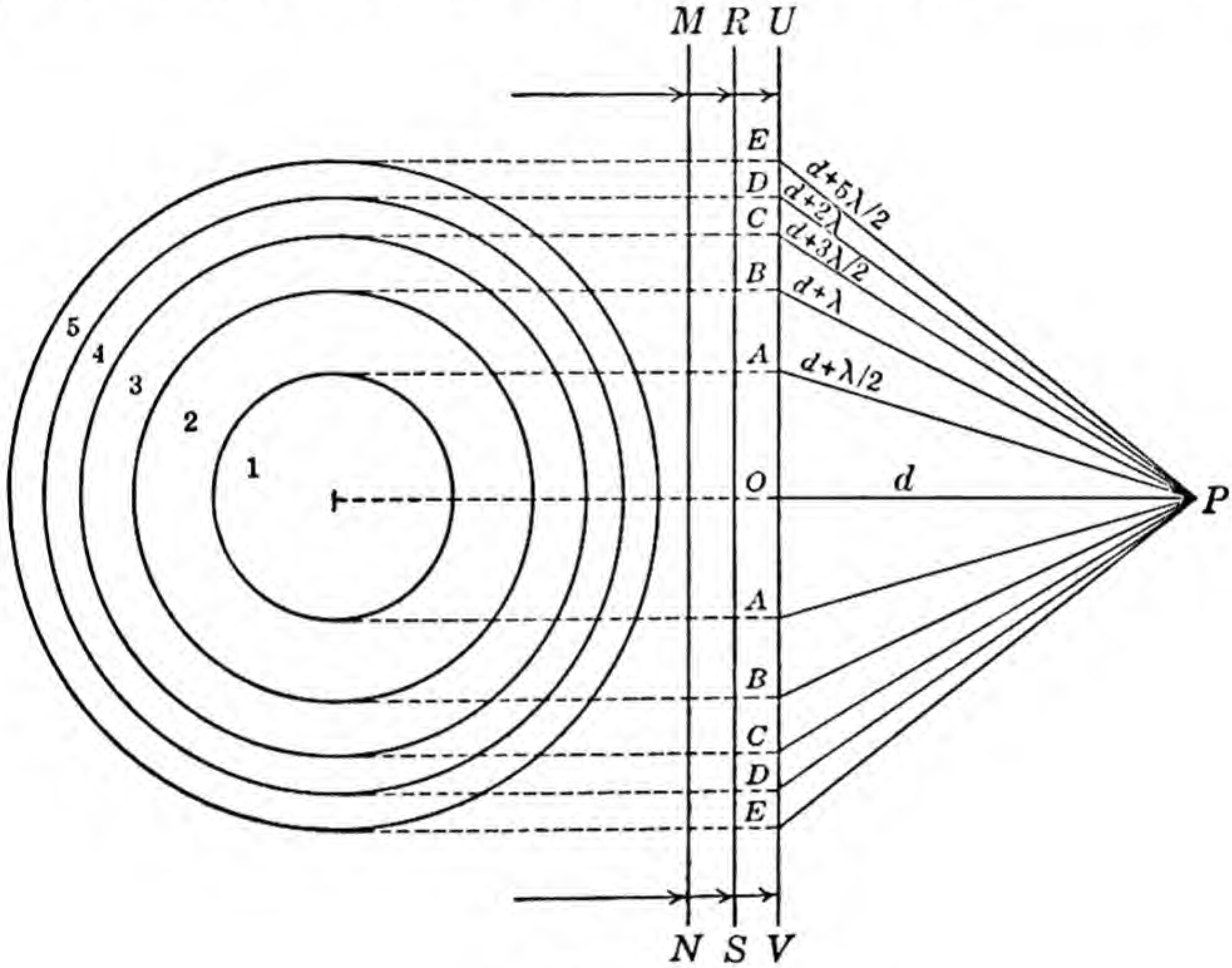


FIG. 204

zero. There will result from these waves an illumination at P. Let the amplitude of the vibration at P due to the first zone be a_1. The second zone in like manner will produce a certain amplitude a_2 at P. Let the amplitudes of the other zones be a_3, a_4, etc.

There are three causes for the difference in magnitudes of the amplitudes produced by various zones. The slightly increasing area of the zones tends to cause the amplitudes to increase. On the other hand, the increasing obliquity of the zones and their increased distances cause them to decrease. Because of these factors the resultant amplitude due to a zone gradually decreases with increasing diameter of the zone.

Since the displacements of the odd-numbered zones are opposite in phase to those of the even-numbered zones, the resultant amplitude at P is

$$A = a_1 - a_2 + a_3 - a_4 + a_5 - a_6 + \cdots \tag{328}$$

In this series, $a_1 > a_2 > a_3$, etc. The difference in any two consecutive values of a is so very small that the amplitude of any zone will be very close to the mean of the two adjacent zones;

$$a_2 = \frac{a_1 + a_3}{2}, \quad a_4 = \frac{a_3 + a_5}{2}, \text{ etc.}$$

Therefore we shall rewrite Eq. 328 as follows:

$$A = \frac{a_1}{2} + \left(\frac{a_1 + a_3}{2} - a_2\right) + \left(\frac{a_3 + a_5}{2} - a_4\right) + \left(\frac{a_5 + a_7}{2} - a_6\right) + \cdots \tag{329}$$

Hence
$$A = \frac{a_1}{2}. \tag{330}$$

So we see that the effect at P of the whole wave front of infinite extent is no more than that produced by half of the first zone. This small illumination is caused by the destructive interference of the Huyghens wavelets.

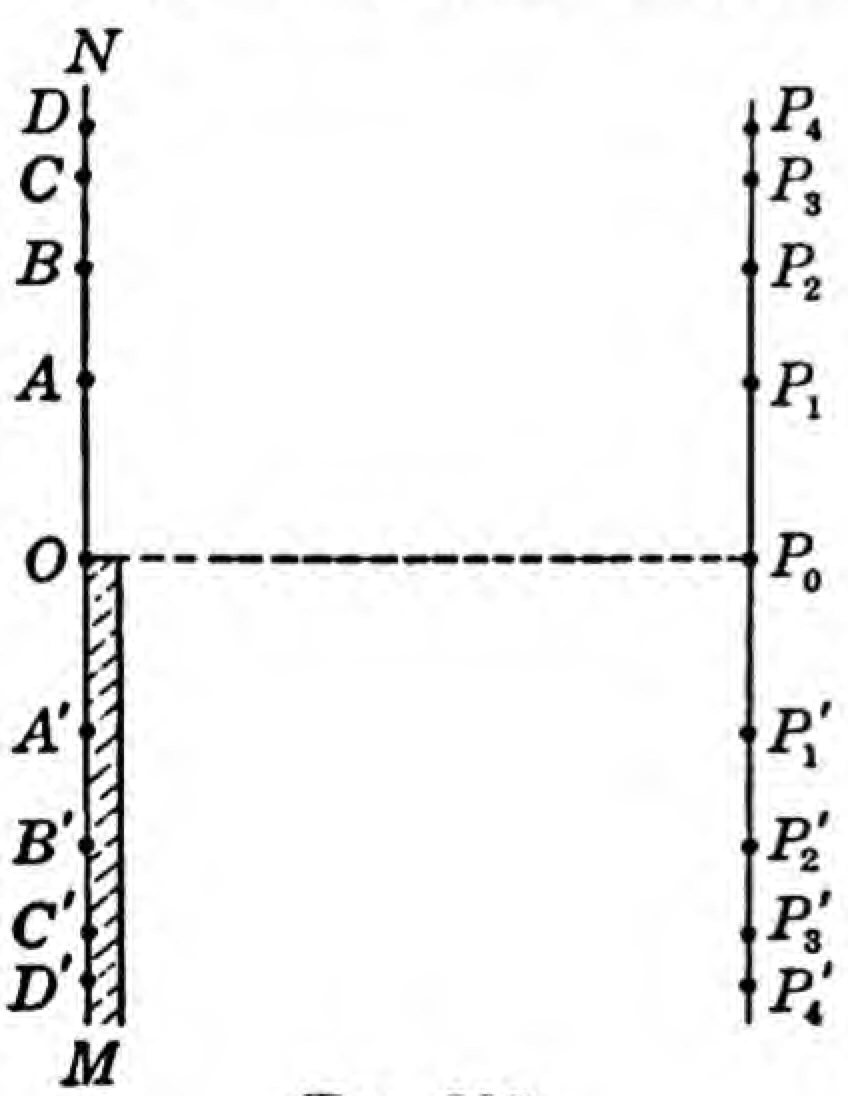


FIG. 205

If we place along UV a plate of glass with opaque rings covering zones 2, 4, 6, etc., then we would have $A = a_1 + a_3 + a_5 + a_7 + \cdots$ which gives a finite sum of considerable size and the intensity of the light P is greatly increased. Such a plate with opaque zones is called a zone plate.

228. The Shadow of a Straight Edge. — Let MN (Fig. 205) be a plane wave front originating from a distant point source. Let an opaque semi-infinite plane OM with a straight edge at O cut off the lower half of the wave front. Consider the line P_0 which is parallel to the edge O and through which all light rays pass which graze by the edge. The first half-period zone for points on P_0 will be an infinitely long strip of width AA' such that

$$AP_0 - OP_0 = \frac{\lambda}{2} = A'P_0 - OP_0.$$

The other zones are indicated in the figure. We shall now examine the resultant amplitude of vibration at points P_0, P_1, P_2, $\cdots$ P_1', P_2', $\cdots$ which are opposite the edges of the half-period zones drawn with respect to the point P_0. Call a_1 the amplitude produced at P_0 by strip OA (equal to that produced by strip OA'); a_2 the amplitude due to strip AB; etc. Just as in the previous section, the resulting amplitude A_0 at P_0 is

$$A_0 = a_1 - a_2 + a_3 - a_4 + \cdots = \frac{a_1}{2}.$$

The amplitude at point P_1 is found as follows. The half-period zones corresponding to the point P_1 are shown in Fig. 206. The point P_1 was selected in Fig. 205 so that its distance to the straight edge was greater than its perpendicular distance to the wave front by $\lambda/2$. So in Fig. 206 one of the lower half-period strips adds to the illumination at P_1. Thus the amplitude produced at P_1 is greater than that at P_0 by the amount a_1.

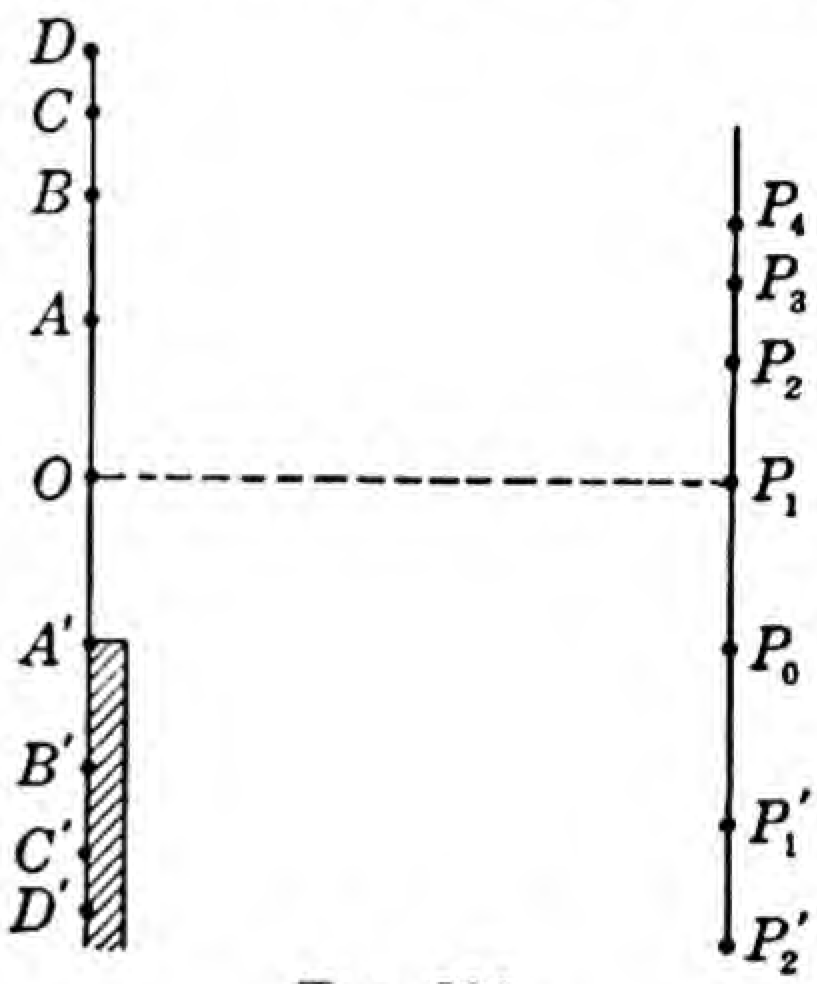


FIG. 206

$$A_1 = a_1 + \frac{a_1}{2} = \frac{3}{2}a_1.$$

So we see that at P_1, just outside the edge of the geometrical shadow, the amplitude is three times greater than at the edge of the shadow, the intensity of illumination being nine times greater.

In a similar fashion, for the position P_2, two of the lower zones are above the edge of the screen and

$$A_2 = \frac{a_1}{2} + a_1 - a_2 = \text{slightly more than } \frac{a_1}{2}.$$

Then at P_3,

$$A_3 = \frac{a_1}{2} + a_1 - a_2 + a_3 = \frac{a_1}{2} + \frac{a_1}{2} + \left(\frac{a_1 + a_3}{2} - a_2\right) + \frac{a_3}{2} = a_1 + \frac{a_3}{2}.$$

At P_4,

$$A_4 = \frac{a_1}{2} + a_1 - a_2 + a_3 - a_4 = \text{slightly more than } \frac{a_1}{2}.$$

At P_5,

$$A_5 = \frac{a_1}{2} + a_1 - a_2 + a_3 - a_4 + a_5 = a_1 + \frac{a_5}{2}.$$

It is thus seen that outside the edge of the geometrical shadow, and parallel to it, a series of bright and dark bands appear which rapidly approach nearer to each other and approach the same intensity (of amplitude a_1).

Within the geometrical shadow at corresponding positions, $P'_1, P'_2, \cdots$

$$A'_1 = a_2 - a_3 + a_4 - a_5 + \cdots = \frac{a_2}{2},$$

$$A'_2 = a_3 - a_4 + a_5 - a_6 + \cdots = \frac{a_3}{2},$$

$$A'_3 = \frac{a_4}{2}, \quad A_4' = \frac{a_5}{2}, \text{ etc.}$$

Thus we see that the intensity continuously falls off toward zero within the geometrical shadow.

The bands outside the geometrical shadow and the gradual approach to blackness within the geometrical shadow may be easily observed and photographed.

PROBLEM

Compute the diameter of the first zone and also the distance P_0P_1 of the first maximum in Fig. 206, for a point 3 meters away from a plane wave front (or from the straight edge) for the sodium D_1 line ($\lambda = 5896 \times 10^{-8}$ cm.). Compute also the same distance for the case of a sound wave of the pitch of A–440 vibrations per second.

229. Diffraction and Rectilinear Propagation of Light. — When the wave theory of light was first proposed, many serious objections were raised. Since it was established by that time that sound was a wave motion, it was believed that if light were a wave motion, then it should show all the phenomena exhibited by sound. Now one of the notable properties of sound is its failure to cast sharp shadows of an opaque body. Sound rays seem to bend greatly around corners. It was thought that light cast perfectly sharp shadows and that, therefore, light could not be a wave motion, As we know now, diffraction occurs in both light and sound. The difference in magnitude of the bending, however, is very large and that great difference is caused by the great difference in wave-lengths in light and sound. In the case of light, as shown by the above sections and problem, the intensity drops to a very low value at only a very small distance inside the geometrical shadow. In sound there is likewise a decrease in the intensity within the

geometrical shadow but the rate of decrease is much smaller than in light.

230. The Diffraction Grating. — The effect of a plane wave passing through a series of narrow parallel slits may be found by investigating the diffraction pattern due to a single slit and then adding up the amplitudes due to each one of a series of equally spaced slits. The following method, although failing to show the effect of increasing the number of slits, gives the result more directly. In Fig. 207, *AB* represents a cross section of a trans-

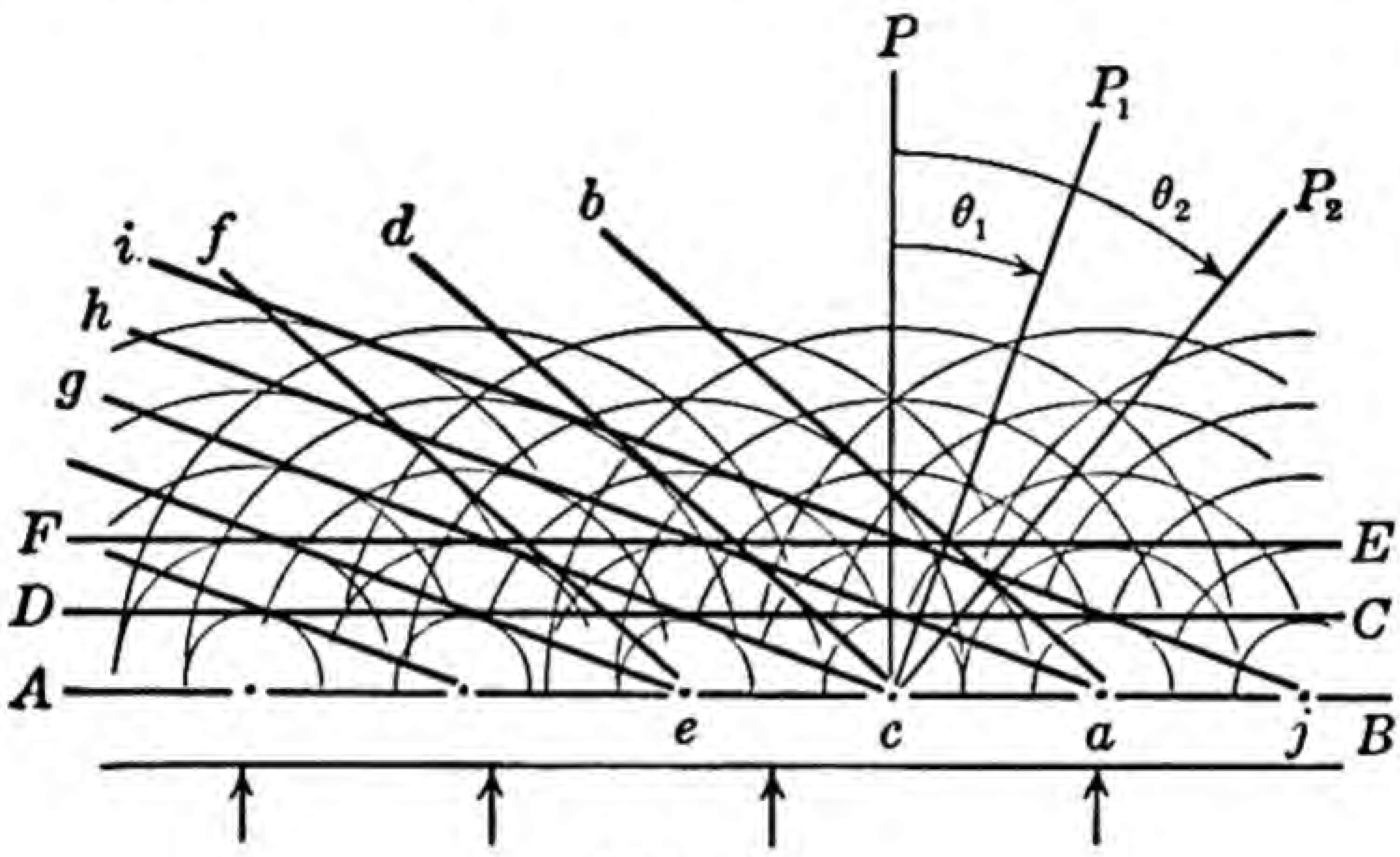


Fig. 207

parent plate. Ruled on the plate with a diamond are a series of parallel scratches which prevent the light from passing through in a regular manner and scatters it in random directions. These scratches are perpendicular to the plane of the diagram. They are represented by the heavy lines along *AB* and for purposes of construction are shown to be much wider than the clear portions of the glass left between the scratches. A series of plane waves approach the grating from below. Part of the regular wave front is scattered by the scratches and the remaining slits become practically line sources of light sending out a series of cylindrical wave fronts. According to Huyghens' principle any tangent plane to a series of waves forms a new wave front. Examining the cylindrical waves in the figure, we see that there are a series of wave fronts which determine planes *CD*, *EF*, etc., which emerge unaffected by the grating. But there is also the series such as *cg*, *ah*, *ji*, etc., whose direction of advance is inclined at the angle θ_1 to the direction of the incident beam. Likewise there are the fronts *ef*, *cd*, *ab*, etc., making an angle θ_2 with the incident direc-

tion. Examination of the triangles whose common hypotenuse is ac gives the following formulae

$$ac \sin \theta_1 = \lambda,$$
$$ac \sin \theta_2 = 2\,\lambda,$$
$$ac \sin \theta_3 = 3\,\lambda.$$

In general, $$n\lambda = s \sin \theta, \qquad (331)$$
where $s = ac$ is called the grating space.

The angle is determined by the wave-length. The student should show that violet light emerges at values of $\theta_1, \theta_2, \cdots$, which are correspondingly less than those of red light. The spectral separation is exactly of the type shown in Fig. 178. If a transmission grating is placed in Fig. 180*b* in place of the prism, the light of each wave-length is brought to a focus and a purer spectrum results. For careful measurements, however, the grating is mounted in place of the prism on a spectrometer as shown in Fig. 181.

It is to be noted that in a spectrum produced by a prism, there is only one spectrum and in that spectrum the violet light is deviated more than the red light. In a spectrum produced by a grating, however, there are many spectra produced and in each one the violet light is deviated less than the red light. The angle at which the light of a given wave-length appears in the various spectra is found by giving n in Eq. 331 the successive values of 1, 2, 3, etc. In the ordinary ruled grating the spectrum nearest the undeviated beam of light (the spectrum of the first order) is the brightest. The succeeding 2nd, 3rd, etc., orders of spectrum grow rapidly weaker in intensity. Each succeeding spectrum is spread out through correspondingly larger angles so that the higher orders of spectra actually overlap.

Polarization and Double Refraction

231. Double Refraction in Calcite. — Calcite is a natural crystalline form of calcium carbonate. It may be broken among cleavage planes into a rhombohedral form as shown in Fig. 208. Each face is a parallelogram with angles of about 101° 53′ and 78° 7′. At two opposite corners, A and G, are three obtuse angles. At all other corners there are two acute angles and one obtuse angle. For a crystal cleaved so that all its edges are the same length, the corners A and G are closer together than any two other

opposite corners. The line AG or any line parallel to it is called the "optic axis" of the crystal. The optic axis is thus any line parallel to the line that is equally inclined to the three natural cleavage faces which form the three obtuse face angles. Any plane which passes through the optic axis and which is perpendicular to a refracting face of the crystal is called a principal plane.

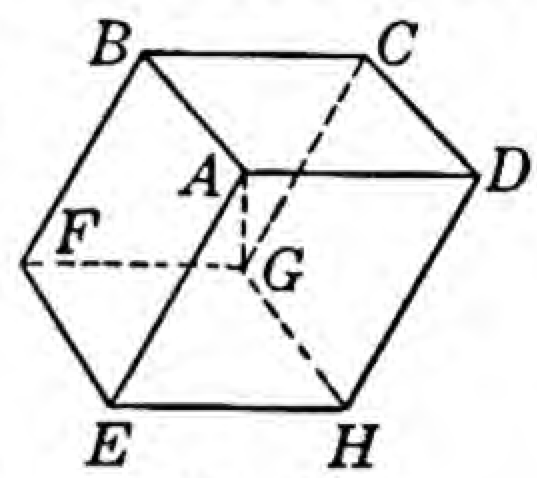


FIG. 208

Calcite has many very peculiar optical properties. The study of these unusual properties has led to the knowledge that in calcite a Huyghens' wavelet, spreading out from a point, consists of two waves, viz. a spherical wave and an ellipsoidal wave. These two waves travel with the same velocity in directions parallel to AG but with different velocities in all other directions. In calcite the ellipsoidal wave is outside of the spherical wave as shown in Fig. 209*a*. In the case of quartz the ellipsoidal wave is inside of the ordinary spherical wave as shown in part (*b*) of the figure.

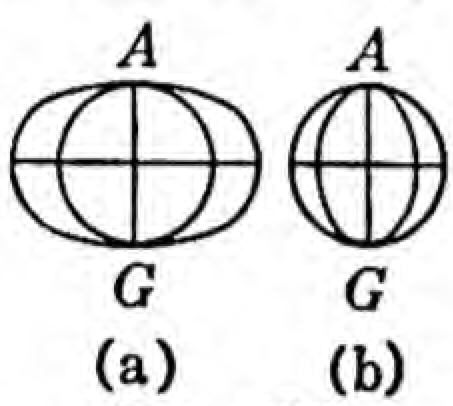


FIG. 209

Crystals such as calcite, quartz, and many others have only one axis along which there is a single velocity. These are called uniaxial crystals. There are many other crystals having two such directions. Such crystals are said to be biaxial.

We shall now show how several important phenomena occurring in calcite are accounted for by these wave forms.

Consider a beam of light with plane wave fronts incident normally on a calcite surface such as in Fig. 210. Let the optic axis AG lie in the plane of the drawing as shown by the dotted lines. At every point N on the wave front there will originate a pair of Huyghens' wavelets. The tangents to the spherical wavelets produce wave fronts parallel to the surface which travel forward along NO and $N'O'$ normal to the surface. But the tangents to the ellipsoidal waves, although also parallel to the surface, travel forward in the directions NP and $N'P'$. Because of their extraordinary behavior, the rays such as NP are called the *extraordinary rays* and the rays such as NO the *ordinary rays*. For the ordinary rays the wave front is always perpendicular to the rays (the direc-

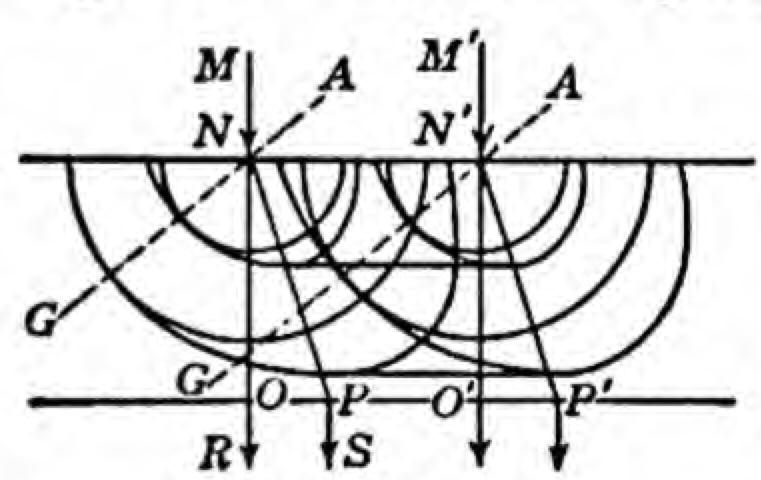


FIG. 210

tion of advance of the wave fronts). For the extraordinary rays the direction of advance of the wave is usually not at right angles to the wave fronts. The extraordinary rays do not obey Snell's law of refraction. In the above case we see that although the angle of incidence of the extraordinary rays is zero, nevertheless there is a definite angle of refraction. Furthermore it can be seen that if the incident ray MN were inclined slightly to the right, the refracted ray NP would approach closer to NO, — thus the peculiar fact that the refracted ray may be on the same side of the normal as the incident ray. In the case shown in Fig. 210 the optic axis lies in the plane of incidence and in that case the refracted ray lies in the same plane. However, when the optic axis does not lie in the plane of incidence, the refracted ray usually is refracted sidewise out of the plane of incidence.

When the ray NO meets the opposite side of the crystal, it will pass through undeviated. The ray NP will be refracted at the second surface so as to emerge perpendicular to the surface. Thus if a point object N is viewed through a calcite crystal, two images of the object will be seen. If the crystal is rotated, the image formed by the extraordinary ray will rotate around the other image. The divergent rays actually observed (similar to Fig. 177) are shown in Fig. 211. Since the ellipsoidal wave lies outside the corresponding spherical wave, it must travel at a greater velocity than the ordinary spherical wave. Hence the index of refraction is larger for the ordinary ray and the image at P_0 is higher than the image P_e. The extraordinary image, when viewed along a normal through a calcite crystal, is always displaced so that it remains in the principal plane passing through the point P.

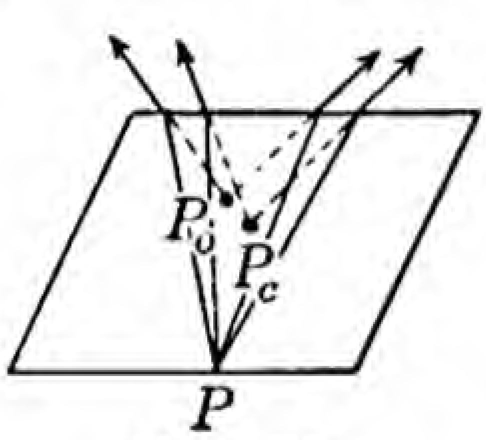


FIG. 211

232. The Two Indices of Refraction of Uniaxial Crystals. — As shown in the above paragraph, the velocity of the ordinary ray is a definite quantity. Let v be the velocity of light in a vacuum and v_0 be the velocity of the spherical wave in a crystal. The index of refraction for the ordinary ray is defined as

$$n_0 = \frac{v}{v_0}.$$

The velocity of the extraordinary wave front varies from the value v_0 along the optic axis to a maximum or a minimum value

according to whether the crystal behaves like calcite or like quartz. Let v_e be this maximum or minimum velocity which is the same along any line at right angles to the optic axis. The ratio

$$n_e = \frac{v}{v_e}$$

is called the index of refraction for the extraordinary ray.

Let MN (Fig. 212) be the upper surface of a calcite crystal cut so that the optic axis lies parallel to the surface. Let a beam of light be incident so that any ray such as LB makes an angle of 90° with the direction of the optic axis. The optic axis in the figure is perpendicular to the plane of the paper. Consider the wave fronts which emerge from a point such as B. Along the optic axis through the point B the wave fronts are coincident and the plane of the paper cuts through the ellipsoids of revolution so as to give circles. While the wave front in air has traveled from A to C, the wave fronts at B have advanced in the crystal to D and D'. For calcite the wave front CD' is that of the extraordinary beam. There are two angles of refraction, r_0 and r_e, and they both lie in the plane of incidence. Applying the method of § 200 and the above definitions, we have

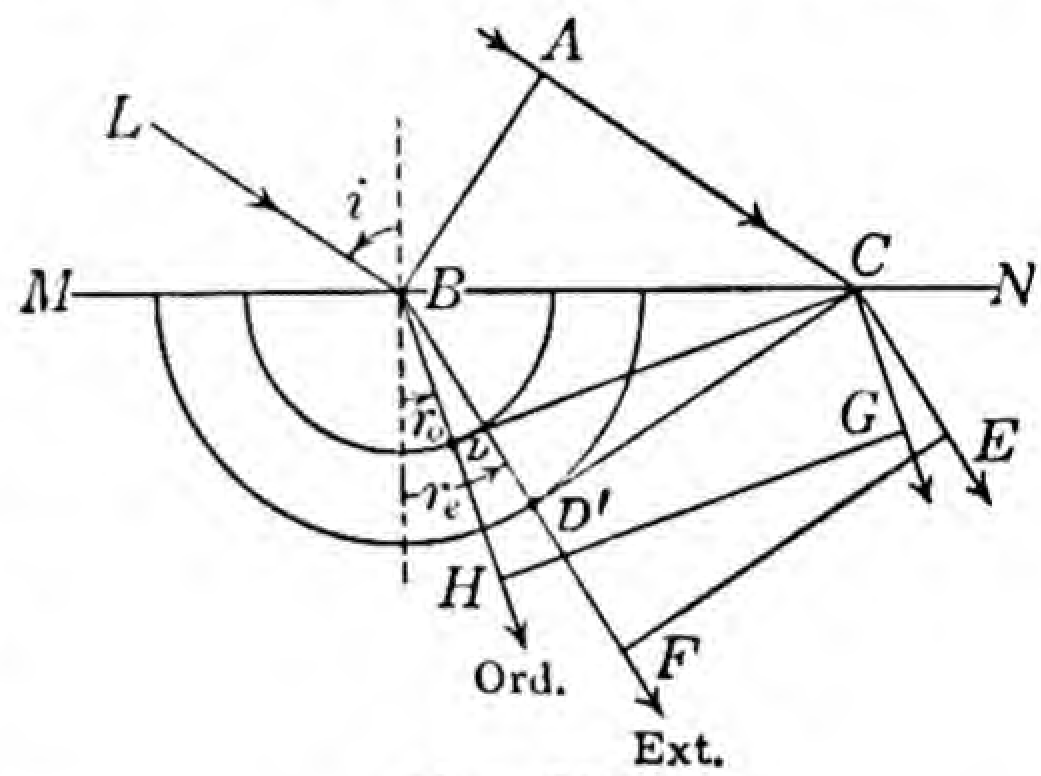


FIG. 212

$$n_0 = \frac{\sin i}{\sin r_0}, \quad n_e = \frac{\sin i}{\sin r_e}.$$

By examining different sections of birefracting crystals it will be found that the above section of a crystal and the above plane of incidence (perpendicular to optic axis) is the only case in which Snell's law will apply.

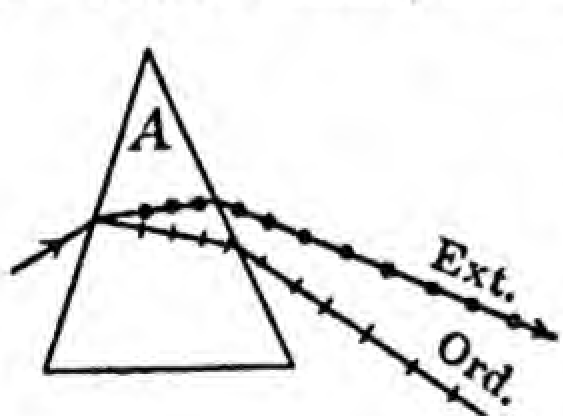


FIG. 213

For actual measurement of the values of n_0 and n_e, a prism is cut from a crystal so that the optic axis is parallel to the refracting edge, (Fig. 213). The prism is rotated until the deviation of first the ordinary and then the extraordinary ray is a minimum and those minimum angles D_0 and D_e measured. After

measuring the face angle A, the values of n_0 and n_e may be computed by Eq. 308.

233. The Nicol Prism. — In order to isolate the extraordinary ray, Nicol modified a calcite crystal as follows. A long crystal (Fig. 214) is cut with its end faces AC and RP equilateral. For comparison, the upper equilateral faces in Fig. 208 and Fig. 214 have been lettered alike. In each figure the optical axis is any line parallel to the line AG and any plane parallel to the plane ACG is a principal plane. The angles ACG and RPA of the original crystal are about 71°.

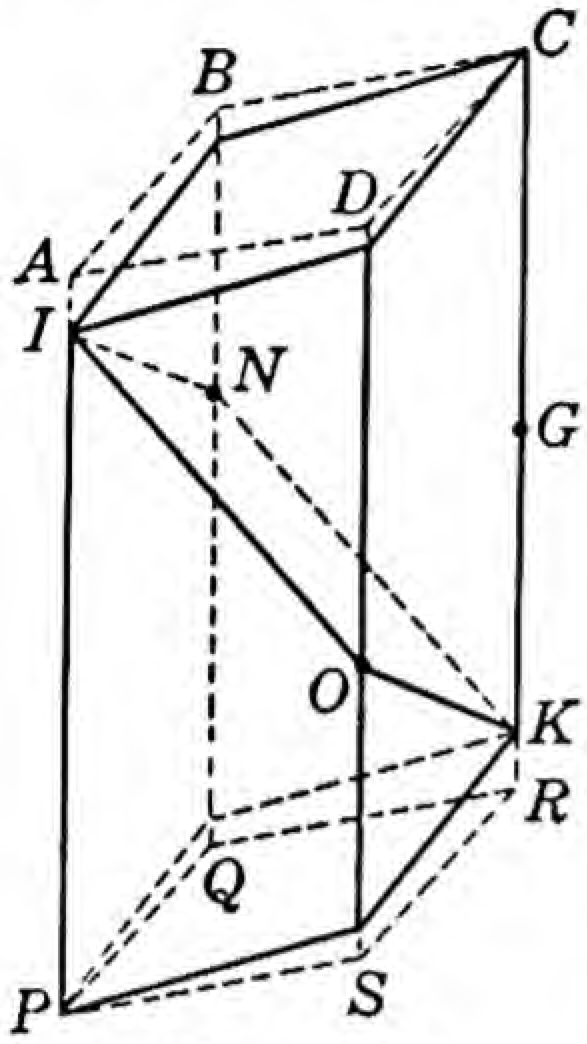


FIG. 214

The first step in making a Nicol prism is that of increasing the slope of the end faces. These faces are cut as shown by the full lines of Fig. 214. The angles ICR and KPA are made about 68°, — a change of about 3°. Next the crystal is cut through the corners I and K along a plane $INKO$ which is perpendicular to the newly formed end faces and perpendicular to the diagonal IC. The two halves of the crystal are then cemented back together with Canada Balsam. Canada Balsam happens to have an index of refraction which is intermediate between the indices of calcite for the ordinary and extraordinary ray. The effect of this is shown in Fig. 215. A ray of light parallel to the long axis of the prism, upon entering the prism, breaks into the ordinary and extraordinary rays. For the ordinary ray the index of refraction of the balsam is less than that of the calcite. The shape of the prism assures that the ray strike the surface at an angle of incidence greater than the critical angle. Therefore, the ray is reflected and is usually absorbed in the blackened sides of the prism. The balsam, however, has an index of refraction for the extraordinary ray which is greater than that of the calcite. So the ray passes through the balsam, being very slightly displaced as in Fig. 178 in passing through the thin layer of balsam. The extraordinary ray emerges from the prism parallel to the incident ray.

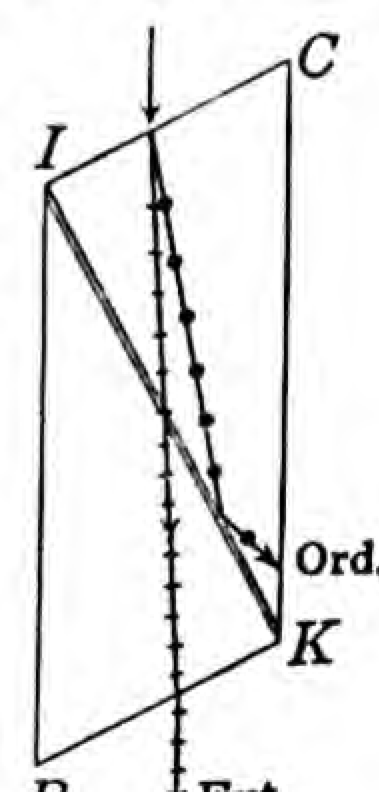


FIG. 215

234. Polarization Accompanying Double Refraction. — Both the ordinary and extraordinary rays behave in unusual ways. If two

Nicol prisms are placed end to end so that the short diagonals, AC, of the faces are parallel, the ray emerging from the first prism passes through the second one. However, if the second prism is turned through ninety degrees the ray which has passed through the first prism is reflected out of the side of the second prism. Viewed from the end of the second Nicol prism no light can be seen. Thus two Nicol prisms "crossed" will transmit no light. It is thus seen that the ordinary and extraordinary rays are alike but have some sort of 90° rotation between them. It is rather difficult to weave together the different phenomena in polarization produced by reflection and refraction and show the gradual growth of our present conception of ordinary light and polarized light. We shall simply state the present conception and show how the explanation fits the facts.

An ordinary unpolarized beam of light consists of magnetic and electric fields varying sinusoidally as mentioned in § 186. The direction of these fields at a given point is continually changing, although, of course, at every instant the electric and magnetic fields are at right angles to each other. When a beam of unpolarized light is incident upon a calcite crystal or any other birefracting crystal, it is found that in each of the two refracted beams the vibrations are only in one plane and the two planes are at right angles. It is customary to talk of the direction of only the electric vibrations because as soon as that is known the direction of the magnetic vibration is determined. It is found that in the ordinary ray the electric field is always oscillating at right angles to the principal plane while in the extraordinary ray the electric oscillations are parallel to the principal plane. In Fig. 215 is shown the cross section of a Nicol prism through a principal plane (parallel to the short diagonal of the face). The vibrations of the electric field in the transmitted extraordinary beam are parallel to the principal plane and parallel to the short diagonal (AC or IC) of the end faces of the Nicol prism. A convenient though meaningless way of remembering those directions is to picture that the electric vibrations in the ordinary ray in Fig. 215 are arrows perpendicular to the plane of the paper. When the ray strikes the reflecting plane, the arrows hit flat and are reflected. In the extraordinary ray the arrows lie as shown by the short cross lines. They hit the plane somewhat slantingly and the arrows penetrate and pass through.

The ordinary and extraordinary rays are thus physically identical to each other except in direction of vibration and the velocity of travel of the rays inside the crystal. However, if the ordinary ray is allowed to pass through the side of the Nicol prism, then the two rays in air have the same velocity and are different in no way except that of direction of vibration. It is then plain why in "crossed" Nicols the extraordinary ray emerging from the first Nicol is reflected out through the side of the second Nicol.

When a Nicol prism is used with unpolarized light incident upon it, it produces an emergent beam which is said to be plane polarized. When so used it is called a *polarizer*. When used to examine the direction of vibration of light polarized by some other means, the Nicol prism is called an *analyzer*.

When a point beneath a calcite crystal, as in Fig. 211, is observed through a Nicol, as can be predicted from the above facts, only the extraordinary image is seen when the short diagonal of the Nicol is parallel to the short diagonal of the upper face of the calcite crystal. When the Nicol is turned through 90°, only the ordinary image is seen. When the Nicol is turned through only 45°, then each image has a component of its vibrations parallel to the short diagonal of the Nicol and both images are seen with the same intensity.

When a Nicol is used as an analyzer to view a beam of plane polarized light, the principle of resolution of vectors shows that the amplitude of the light transmitted through the analyzer is given by the expression $A = A_0 \cos \theta$, where A_0 is the maximum amplitude and θ is the angle measured from the position of the analyzer when the maximum intensity is observed. Therefore the intensity of the transmitted beam varies as the square of the cosine of the angle θ.

235. Polarization by Reflection. — When light is incident on a plane sheet of glass (Fig. 216*a*) about five per cent of the light is reflected from the front face. The remainder is refracted and passes to the back surface where about five per cent is again reflected. From both surfaces somewhat less than ten per cent is reflected and the remaining amount is transmitted. When the reflected light is examined through a Nicol prism, it is found to be strongly although not completely polarized. When the Nicol is placed with its short diagonal in the plane of incidence, there is but little light transmitted. When turned through ninety degrees,

the light is of maximum intensity. Thus, in the reflected rays the electric vibrations are predominately perpendicular to the plane of incidence. As the angle of incidence is changed there is found to be one angle, called the *angle of polarization* or *the polarizing angle*, at which the polarization is the greatest, *i.e.* the light intensity is nearest to zero when the Nicol has its short diagonal parallel to the plane of incidence. For certain substances the polarization is nearly complete at the polarizing angle. By comparing Figs. 216*a* and 215, it will be seen that the same scheme can be used for telling by rote the direction of the vibration of the electric field in the reflected beam. Of course, the process in the two cases is quite different. There is a great difference in the composition of the transmitted beams. In the Nicol prism the transmitted beam is completely polarized. When the ray transmitted through an inclined sheet of glass is examined through a Nicol prism, it is found that the light intensity is only slightly decreased as the short diagonal of the end face of the Nicol is turned perpendicular to the plane of incidence. That fact shows that at a single reflection only a portion of the component of the electric field parallel to the surface (perpendicular to the plane of incidence) is reflected. The following argument should make the above facts coherent. In the unpolarized incident beam in which the vibrations are continually changing direction, we may consider the components of the vibrations to be equal along any two mutually perpendicular directions. The energy associated with the vibrations must then be equal. Therefore, if all the vibrations along one axis were removed, the remaining beam would be reduced to half the intensity of the original beam. Hence, in Fig. 216*a*, if the reflected ray were completely polarized and contained half of the energy of the

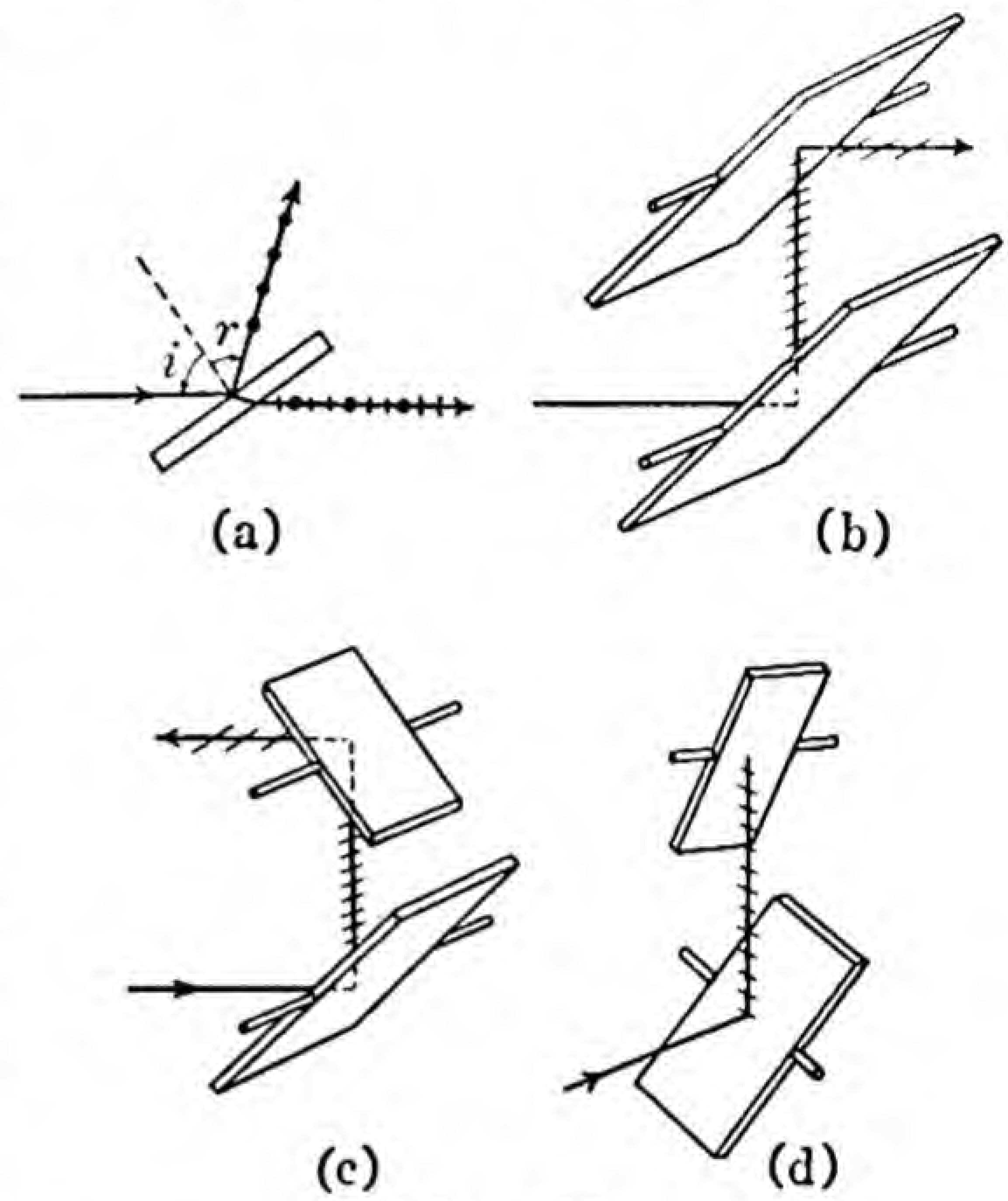


FIG. 216

incident beam, the refracted beam would have to be also completely polarized. On the contrary, we know that the reflected ray, even when completely polarized, has less than ten per cent of the energy of the original beam. Then the remainder, forty per cent of the vibrations perpendicular to the plane of incidence, must be transmitted and is mixed with those vibrations parallel to the plane of incidence which make up the other fifty per cent of the energy. Thus, for the case of a ten per cent reflected, completely polarized beam the intensity of the transmitted light should show a 50 : 40 ratio of intensity when viewed through a Nicol prism with its short diagonal first parallel and then perpendicular to the plane of incidence. In this discussion we have ignored the energy lost in transmission through the glass.

In order to get greater intensities of light polarized by reflection, a half dozen or more plates may be placed on top of each other. At each successive plate a certain per cent of one of the components of vibration is reflected until nearly all of the energy of that component is either reflected or absorbed. Thus at each successive plate the transmitted beam is freed of more of that component of vibration and becomes a more completely polarized beam. A half dozen plates assure moderately complete polarization of both beams, although not approaching the degree of completeness of polarization of the two beams produced by double refraction.

A simple polarimeter for study of reflection may be made by mounting two glass plates on parallel axes as shown in Fig. 216*b*. The lower plate is tilted to the polarizing angle. If the upper plate is tilted parallel to it, then the light reflected from the lower plate will be reflected from the upper plate, as shown in parts (*b*) and (*c*) of the figure. The direction of vibration of the electric field, as indicated, is perpendicular to the plane of incidence. In part (*d*) of the figure the upper plate is turned so that its axis is perpendicular to the axis of the lower plate. Let a beam of light be incident in a plane perpendicular to the axis of the lower plate. The reflected portion has its electric vibrations as shown. When this reflected ray hits the second plate, practically all the light will be transmitted and if the upper plate is of very black glass, this transmitted beam is practically all absorbed. Hence very little light is either reflected or transmitted by the second plate.

CHAPTER VI

ELECTRON PHYSICS AND THE QUANTUM THEORY

The following discussion is intended as a brief survey of some of the fundamental concepts of the theory of electrons and of the experimental basis of our knowledge in this field. It affords a glance at the methods employed in studying electrons.

Electrons and Their Properties

236. Cathode Rays. — If a glass tube (Fig. 217), into the ends of which are sealed metal electrodes connected to a source of high potential such as a static machine or an induction coil, is connected to an exhaust pump and exhausted, a peculiar striated luminescence is set up throughout the length of the tube. The color and character of this luminescence depends upon the kind of residual gas in the tube and upon the degree of exhaustion. With a sufficiently high vacuum, the striae disappear and a greenish fluorescence appears on the sides of the tube.

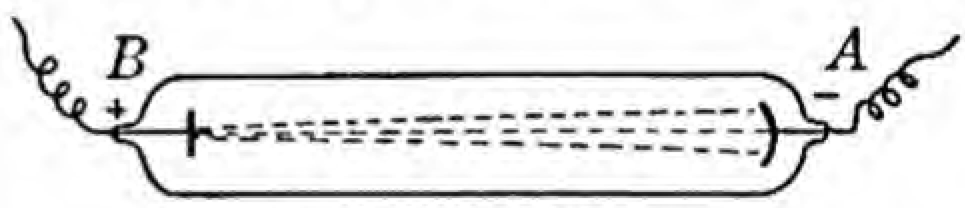


Fig. 217

J. J. Thomson in 1897 showed that the greenish fluorescence was produced by negatively charged particles shot off from the cathode with high velocity. With a tube such as is shown in Fig. 218, a portion of the discharge passes from the negative terminal A through the hole in the anode B into the enlarged portion of the tube and produces a fluorescence on the opposite wall of the bulb at D. This cathode stream, as it is called, may be deflected by a magnet and caused to strike the electrode at C. If the electrode at C is connected to an electroscope, an accumulation of negative charge will be indicated whenever the ray is deflected so as to strike it. The stream is deflected by a magnetic field always in such a direction as to indicate a flow of negative particles out from the cathode A (see § 138).

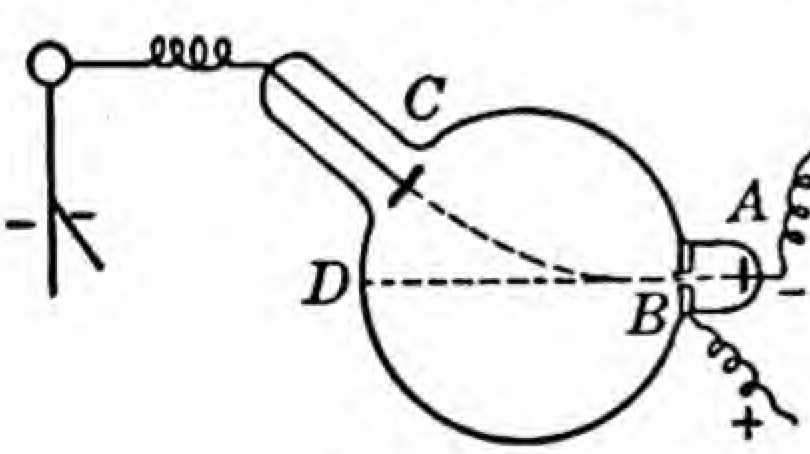


Fig. 218

237. Measurement of the Velocity of Cathode Ray Particles. — If a tube is constructed of the form shown in Fig. 219, a cathode stream emerging through the aligned holes aa' passes out into the enlarged portion to D. Condenser plates at CC may be connected to a battery and an electric field E established across the

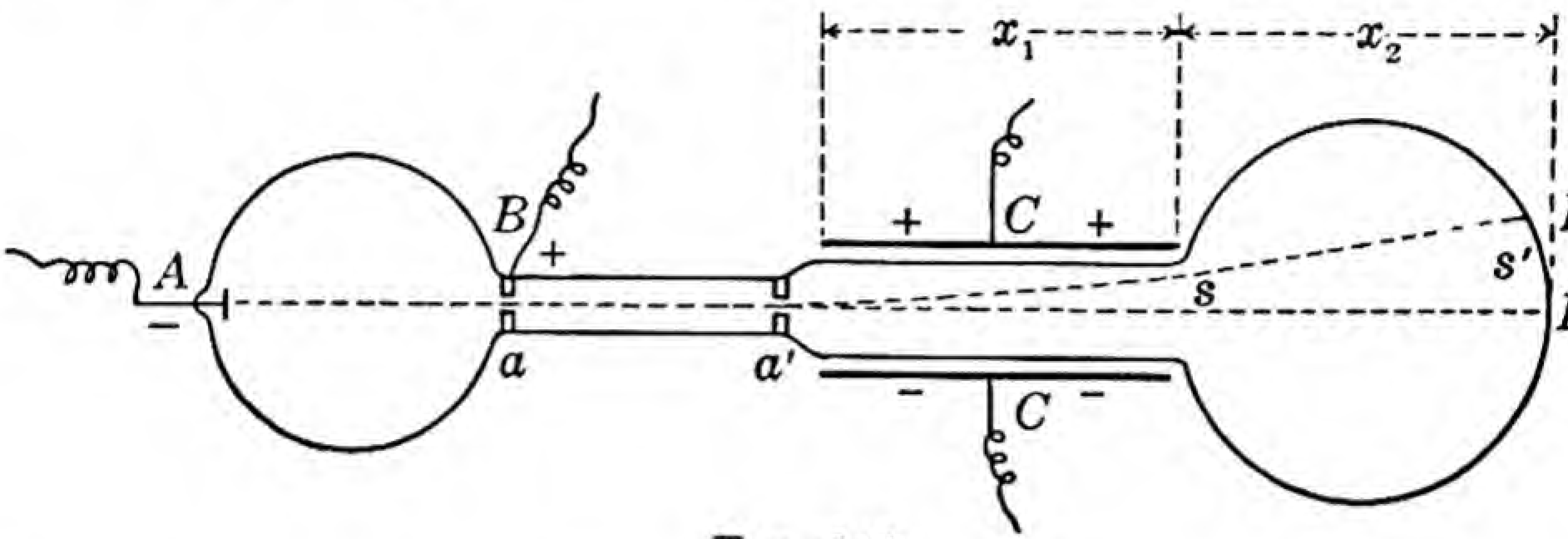


FIG. 219

tube. This electric field will cause a deflection of the moving negative particles. The deflecting force is Ee, where E is the electric field and e the charge on the negative particle or electron.

If a magnetic field is applied at right angle to the electric field and in the same section of the tube, a deflection of the ray will also be produced by this means. If the cathode stream is made up of negative particles moving with high velocity, that stream is equivalent to an electric current as shown by Rowland (§ 174) and will be acted upon by a magnetic field just as is a wire carrying current. The side push on a wire carrying a steady electric current under the above conditions is Hil (Eq. 211). It is evident that ev is equivalent to il because $il = Ql/t = Qv$. Therefore the side push on each electron in the cathode ray stream is Hev.

The magnetic field is set up at right angles to the electric field and its value so adjusted as to counter-act or annul the deflection due to the electric field, in which case there will be no deflection of the fluorescent spot. The electric and magnetic forces are then equal and we may write $Hev = Ee$. From this we may obtain the velocity, thus,

$$v = \frac{E}{H}. \tag{332}$$

238. X-Rays. — In 1895 Röntgen in studying the cathode ray discharge noticed that certain bodies placed near such tubes were excited to fluorescence and that photographic plates even though thoroughly protected from light were blackened. The radiations producing these effects were called X-rays. X-rays are now

known to be of the nature of light of extremely short wave-length. The construction of an ordinary X-ray tube is shown in Fig. 220. The cathode consists of tungsten filament heated to incandescence. Electrons ejected from the hot filament are attracted to

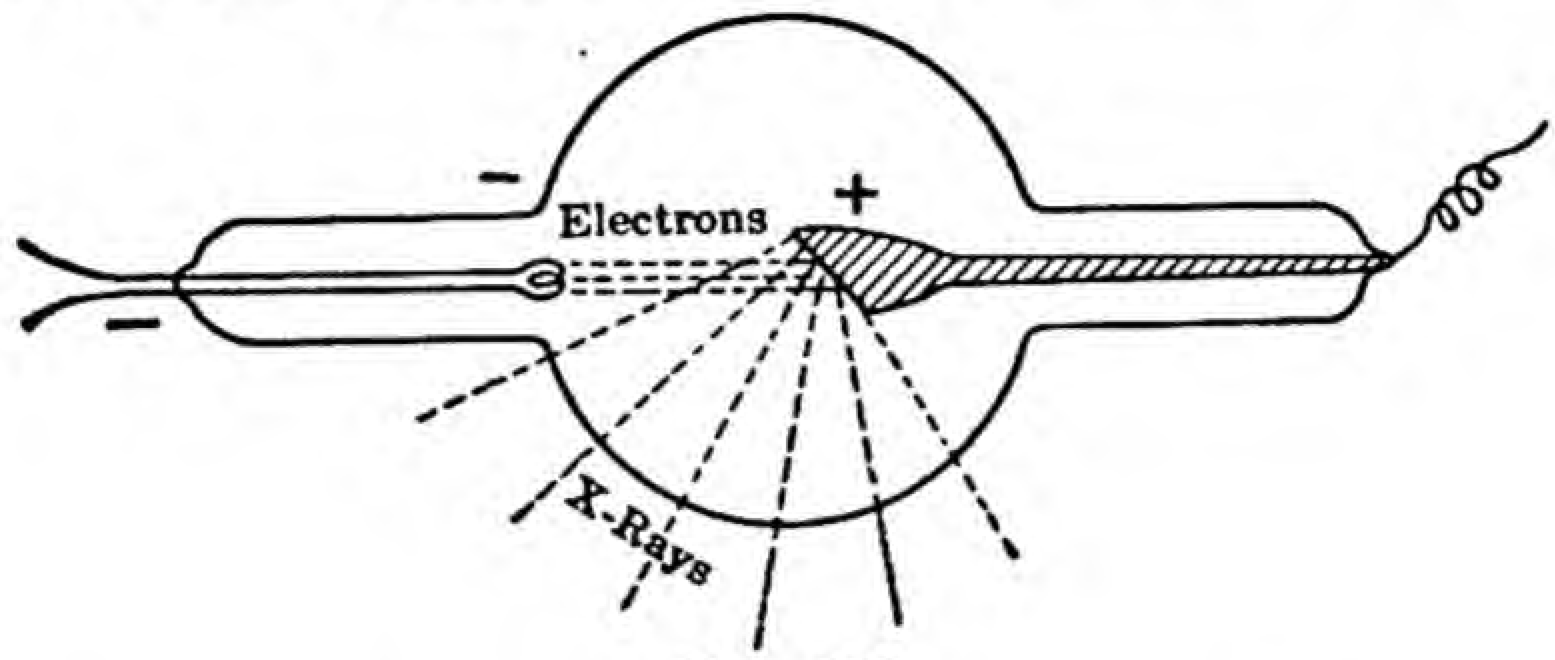


FIG. 220

the anode, usually a heavy block of tungsten. X-rays are emitted in all directions from the face of the anode.

239. Ionization by X-Rays. — Besides the property of penetrating materials and producing shadow pictures on a screen or photographic plate, X-rays have the property of rendering gases conducting. The absorption of the energy of the X-rays by a molecule causes the ejection of an electron, thus leaving the molecule electrically positive. In this state the molecule is called an *ion*. We have thus produced in the gas positive ions and negative particles or electrons.

If we have two plates charged as shown (Fig. 221), one grounded, the other connected to an electroscope, and direct a beam of X-rays between them, the electroscope immediately discharges, showing that the gas between the plates is no longer insulating. The X-rays have ionized the gas, thus rendering it conducting and causing the plates to discharge.

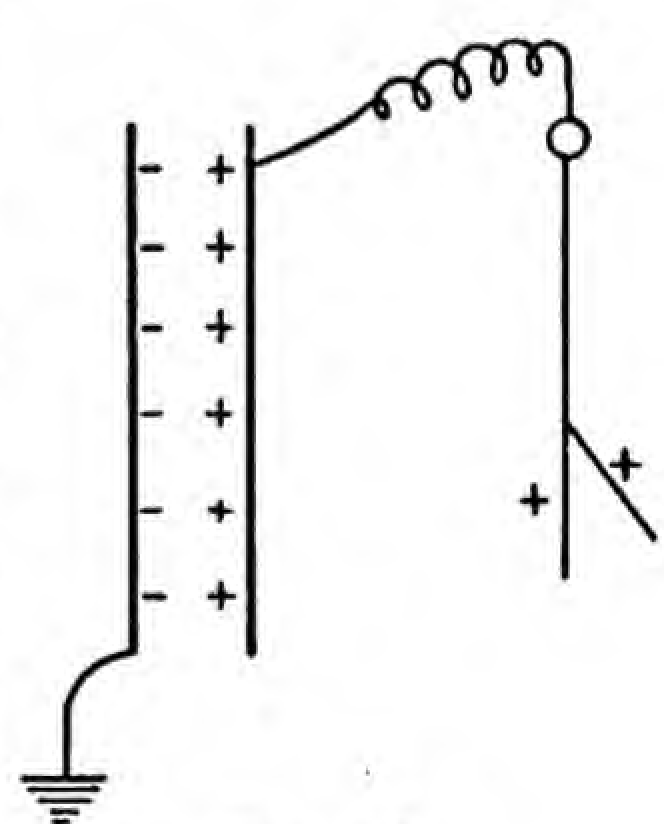


FIG. 221

240. Radioactivity. — The study of X-rays and of the fluorescence produced by X-rays in certain salts led to the discovery by Becquerel in 1896 that salts of uranium produced the same effects as X-rays. Mme. Curie discovered that thorium and its compounds emitted X-rays or were "radioactive." Her search led to the discovery in pitchblende of the far more active and hitherto unknown element, radium.

Radioactive materials are now known to emit radiations of three different sorts, the so-called α-, β-, and γ-rays. The γ-rays are of the same nature as X-rays although more penetrating, being of even shorter wave-length. The β-rays are streams of negative electrons having considerably higher velocity than the cathode ray particles, although in all other respects similar. The α-rays are composed of positively charged particles of very much greater mass than the negative particles. The α particles have been thoroughly identified as the nuclei of helium atoms.

241. Determination of the Value of the Ratio e/m. — Referring to the cathode ray experiment (Fig. 219), we note that the electron while between the charged plates experiences a constant force, Ee, due to the electric field, and is deflected from its straight path just as a body shot horizontally is deflected by gravity. It is displaced an amount s at the edge of the plates and an amount s' at the end of the tube. The horizontal distance the electron travels while between the charged plates is $x = v_x t$, where v_x is the horizontal velocity of the electron and t is the time it is in the region between the plates. For the distance s that the body is deflected by the field, we have $s = at^2/2$. Eliminating t, we have

$$s = \frac{ax_1^2}{2\,v_x^2} \quad \text{or} \quad v_x^2 = \frac{ax_1^2}{2\,s}.$$

Now $a = f/m = Ee/m$; hence,

$$\frac{e}{m} = \frac{2\,sv_x^2}{Ex_1^2}. \tag{333}$$

s may be evaluated in terms of s' and x_2, the horizontal distance traveled after leaving the electric field.*

From Eq. 333 we can determine the value of the ratio e/m for any charged particle. For β-ray particles (electrons) the value is about 1760×10^4 e.m.u. gm.$^{-1}$.

For the α- and β-rays from radioactive material the value of e/m and of v can be obtained by essentially the same method,

* Let the electrons travel the length of the plates x_1 in the time t_1 and the distance from the plates to the end of the tube, x_2, in time t_2. Consider the x axis parallel to the plates and the y axis perpendicular to them. Then we have the equations:

$$s' - s = v_y \cdot t_2 = a_y t_1 \cdot t_2 = \frac{Ee}{m}\frac{x_1}{v_x} \cdot \frac{x_2}{v_x} = \frac{Eex_1x_2}{mv_x^2}.$$

If this is substituted in Eq. 333 and E/H for v_x, the expression for e/m is obtained in quantities which can all be measured.

except that instead of a cathode at A (Fig. 219) we have a small amount of radioactive material. The stream passing through to C is separated as it passes through the electric or the magnetic field into two rays which are deflected in opposite directions, indicating the α-rays to be composed of positive particles. The value of e/m for α-particles is about 4.8×10^3. These particles are found to possess two of the elementary + charges, that is, they lack two electrons.

The value of e/m found for hydrogen ions by electrolysis is 9.6×10^3 and since the value of e/m for the electron is 1760×10^4, the mass of the electron appears to be about 1/1840 that of the hydrogen atom. The mass of the α-particles appears from these data to to be about four times that of the hydrogen atom.

242. Determination of the Value of *e*. — Although the value of the ratio e/m may be obtained as above described, the separate values of e and m are not so readily measured. C. T. R. Wilson in 1897 made use of the fact that in a chamber containing saturated water vapor condensation occurred about charged particles as nuclei. By exposing the air in such a chamber to X-rays, thereby ionizing the air and then producing a sudden small expansion, condensation would occur on the negative ions only, if the expansion (and resultant cooling) did not exceed a certain critical value. By observing the rate at which the droplets fell under gravity, the size of the droplets could be computed. The total amount of water precipitated could be computed from the temperature drop and the total charge measured by attracting the charged droplets to condenser plates in the chamber and noting an electrometer deflection. The method was later improved by H. A. Wilson by placing a plate in the top of the chamber which was positively charged while the water surface below was made negative. The electric field was adjusted so as just to arrest the downward motion of the cloud. Thus for each individual droplet the balanced forces are expressed by the equation $Ee = Mg$ where E is the electric field, M the mass of the droplet, e the charge upon which the condensation occurs, g the acceleration of gravity. M is determined by recourse to Stokes' law, which expresses the relation between the velocity with which the droplets fall through the air and their radii. This law is $v = 2\,gr^2/9\,\mu$, where g is the acceleration of gravity and μ is a constant for a given gas or liquid (called the coefficient of viscosity). The value of μ is a measure of the

viscous resistance to the movement of a body through a liquid. The values of e obtained by this method were in the neighborhood of 4.0×10^{-10} with a considerable range of variation.

R. A. Millikan modified the above method, using a single charged oil droplet suspended in space against gravity by the pull of an electric field. The radius of the droplet was determined by observing the rate of fall when the field was released and applying Stokes' law. The size of the charge was determined by the strength of field necessary to balance gravity. The equations are the same as above. By this method the charge on a droplet was found always to be some exact multiple of 4.770×10^{-10} e.s.u.

From kinetic theory data the number n of atoms in a gram of hydrogen has been calculated as about 6.064×10^{23}. In electrolysis we find that 96490 coulombs will deposit a gram equivalent of any metal. Hence, the charge per ion of hydrogen is $96490/n = 1.591 \times 10^{-9}$ coulombs or 4.773×10^{-10} e.s.u., which agrees with the value found above for the electron.

Thermionic Emission and Vacuum Tubes

243. Introduction. — For nearly a century it has been known that hot bodies do not retain a charge and that cold bodies lose their charge due to the presence of a hot body nearby. In 1880 Elster and Geitel made a careful investigation of these phenomena. They found that when a filament placed in a vacuum became white hot due to an electric current sent through it, negative electricity was given off and would pass to a neighboring electrode. Edison, in 1883, noted the same phenomenon. In 1899 J. J. Thomson measured the value of e/m for the negative particles given off by hot bodies and found it to be exactly the same as for electrons. All other evidence tends to prove that the particles are electrons. This phenomenon of *the emission of electrons from hot bodies has been called the Edison Effect.*

In metallic bodies there are many free electrons and they move about with random velocities determined by the temperature of the body. As the temperature of the body is raised the electrons gain energy and their average velocities increase. Due to collisions the individual electrons interchange energy. Suppose that an electron near the surface is collided with successively by several neighboring high-speed electrons and is thus given a very high velocity. If this velocity is directed away from the surface,

the electron may actually be shot through the surface against the forces holding it in the metal. We might expect this to happen occasionally, though rarely, even at room temperature. However, as the temperature of the metal is raised, all the electrons have greater average velocities and there will be more electrons which attain velocities sufficient for escape. Any electron which possesses more energy than that necessary to just escape through the surface will be ejected with a finite velocity. Thus a few electrons escape with little or no velocity, a few with high velocity, and the majority with various moderate velocities determined by the temperature.

The rate at which electrons are emitted from a given area of a substance at a given temperature depends greatly on the surface material. Pure platinum or tungsten emits relatively few electrons compared to the same metals when coated with small amounts of barium, thorium, or other substances.

For a body of given material, the rate of emission of electrons depends upon the temperature of the body. The number of electrons emitted per second is given by the formula:

$$N = AT^2 e^{-\frac{b}{T}}, \tag{334}$$

where T is the absolute temperature and A and b are constants depending on the material of the emitting body.

If the emitting body is insulated, it will become positively charged by the loss of electrons. The cloud of negative electrons around the body will then be attracted back toward the positively charged body. As the total number of emitted electrons becomes larger, the emitter becomes more positive and the attractive forces increase. Equilibrium will be reached when as many electrons are being pulled back into the body as there are electrons leaving the body. If a plate P (Fig. 222) is put near the emitting body F and is maintained at a positive potential with respect to it by a battery B, then a continuous current of electrons will flow from the filament to the plate, through the battery and back to the filament.

244. The Diode. — A vacuum tube containing a plate and a filament which may be heated is called a two-element tube or diode (two electrodes). A diode is usually connected as shown in Fig. 222. The battery A supplies current to a filament F, the temperature of the filament being regulated by the rheostat R. The plate P is made positive with respect to the filament by the battery B.

As the electrons are emitted from the filament, they are attracted to the positive plate P, pass through the milliammeter, m.a., through the battery, and back to the filament. In a metallic conductor only electrons can move. Since, however, the direction of a current was very early chosen as that in which the positive charges would flow, we say that the current flows in the opposite direction to that of the electrons. Thus the current flows out of the B battery, up through the ammeter, and counter-clockwise around the circuit.

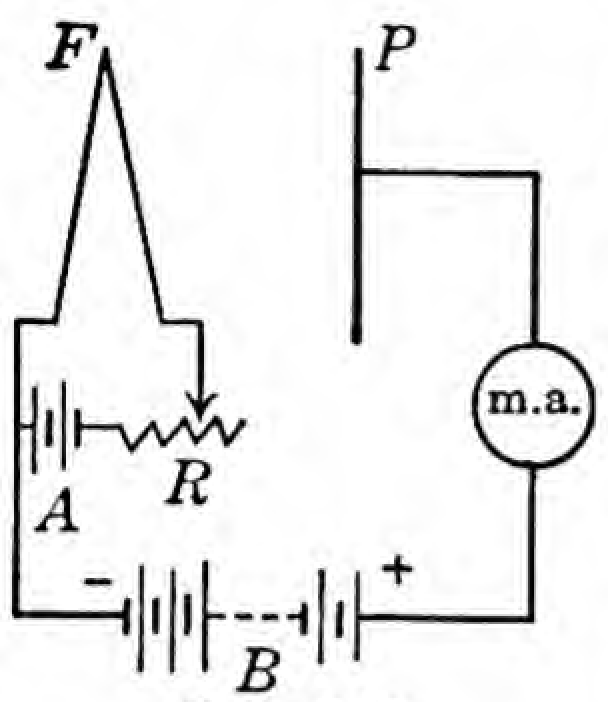


FIG. 222

The electron current as it flows through the space *from the filament to the plate is called the plate current.* The magnitude of this current depends on the temperature of the filament and potential of the B battery. To show the relation between the current I_p, the filament temperature T, and the plate potential E_p, we may plot the values of the current along the y-axis, the temperature along the x-axis, and the potential along the z-axis. A warped surface results. Various plane cross-sections parallel to the x-y plane may be represented on a single diagram such as Fig. 223. Sections parallel to the y-z plane are shown in Fig. 224.

Let the plate potential be maintained constant. If every electron that is emitted could reach the positively charged plate, then we should expect that the plate current would be represented by Eq. 334. This relation is shown by the dotted curve in Fig. 223. We shall see that in practice this curve may be approached only when E_p is very large. When E_p is small, say E_1, I_p varies as shown by the lower curve in the figure. The plate potential being small, the electric field between the filament and the plate is weak. In the space between the filament and the plate there are a considerable number of electrons in transit. These electrons make up what is called the *space charge.* This space charge repels the electrons just emitted and forces some of them back into the filament, thus decreasing the current. When T is small,

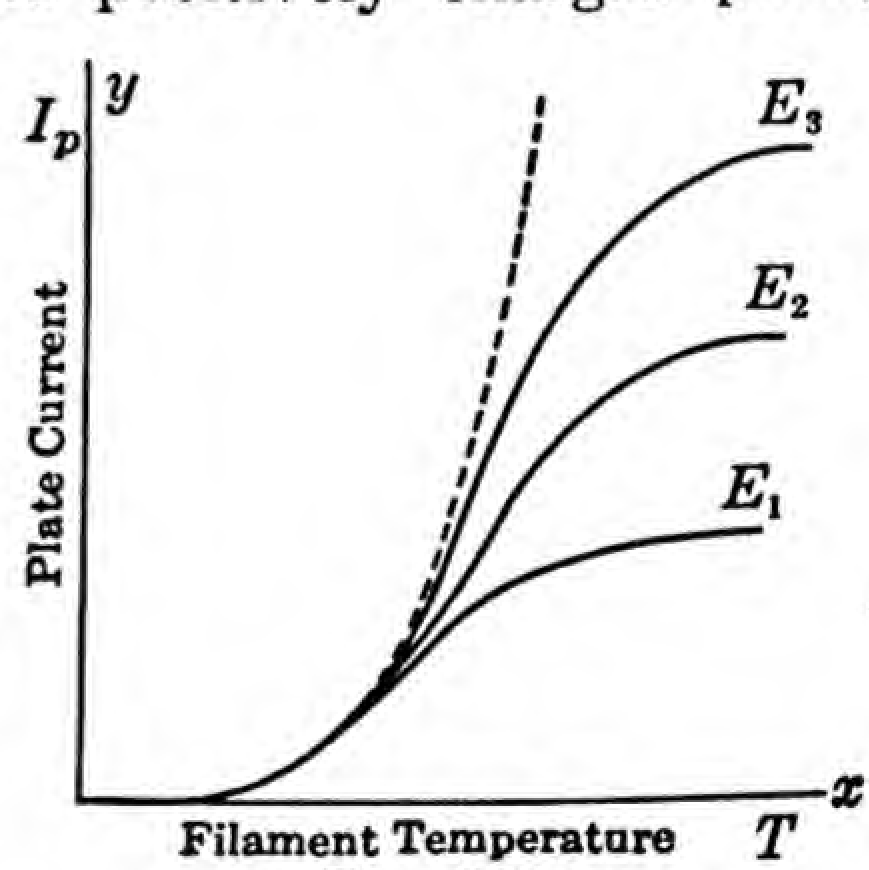


FIG. 223

there are only a few electrons emitted and hence a small space charge and very little retarding effect. As the filament temperature is raised the electrons are emitted at a much greater rate, the number in transit (the space charge) increases, and the retarding effect becomes larger. As shown in the figure, the increasing space charge causes the E_1 curve to leave the dotted line more and more. Finally the space charge gets so large that it prevents any further increase in I_p. This is called a condition of *saturation* and the maximum value of I is called the *saturation current.* Now when the plate potential is raised to E_2, the electrons are speeded up and the space charge is decreased. The value of I_p is increased for all values of T and a larger saturation current is obtained. The equation, represented by the dotted curve, is derived on the assumption that there is no space charge limitation and that the temperature of the filament is the only limiting factor. Actually, however, a deviation from the dotted curve is caused by the space charge limitation.

Now let us consider how I_p changes if T is kept constant and the plate potential E_p changed. The results are shown in Fig. 224. Consider first that the filament is at some low temperature T_1 such that the rate of emission of electrons is relatively small. When E_p is small, the electrons move slowly across the plate, the space charge effect is large, and therefore I_p is small. As E_p increases I_p increases. Since the filament emission is not large, a rather low value of E_p will suffice to reduce the space charge practically to zero. When that condition is reached, we have seen in the previous paragraph that all the electrons pass across to the plate as fast as they are emitted. The current cannot be increased beyond this unless the rate of emission from the filament is increased by raising its temperature. For a higher temperature T_2 the rate of emission is larger and therefore a higher voltage E_p must be applied before the space charge becomes effectively zero. The saturation current now is greater than before. The saturation value in each case is given by Eq. 334. Theory shows that if the velocity

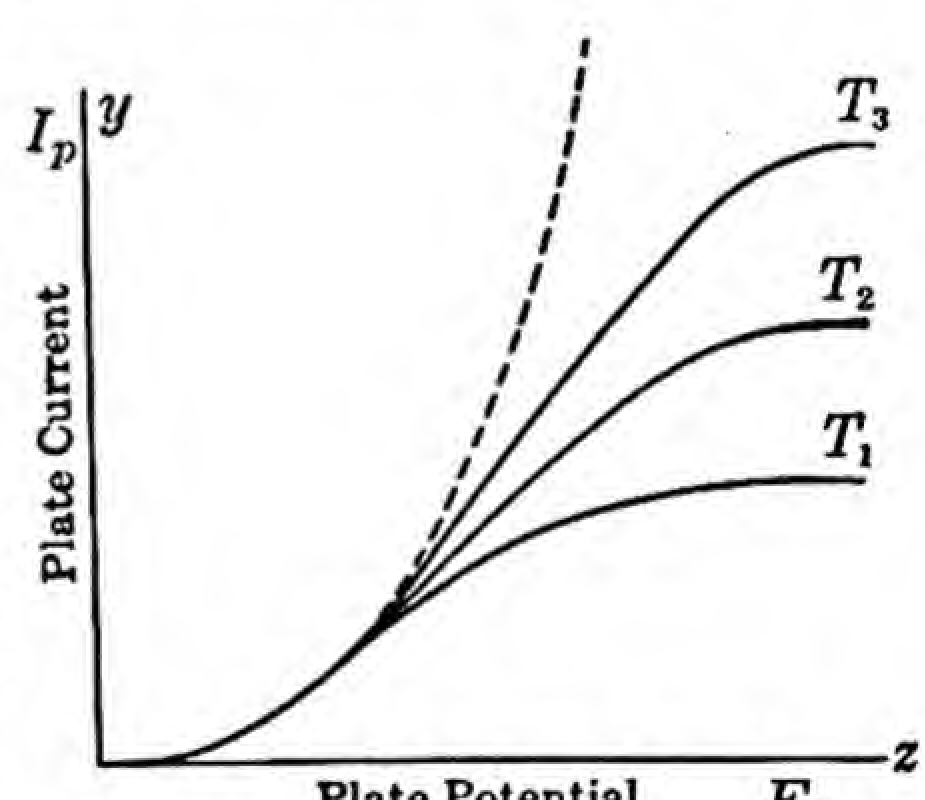


FIG. 224

with which the electrons are emitted from the surface is small, then the following relation holds.

$$I_p = CE_p^{\frac{3}{2}} \tag{335}$$

where the value of C depends on the size, shape, and separation of the electrodes. The dotted line in Fig. 224 is the graph of the equation. The equation is obtained under the assumption that there is no limitation on the supply of electrons and that the space charge is the only limiting factor. Actually, however, there is a deviation from the dotted curve caused by the temperature limitation on the rate of emission.

Although the curves in Figs. 223 and 224 appear very similar, they are quite different. The former set of curves emerges from the origin and rises rapidly as an exponential function of T before saturation effects begin to be noticeable. The latter set rise only as the $\frac{3}{2}$ power of E_p. The student should note carefully the causes of the rise in each case and the factors which cause saturation.

If the B battery is replaced by an alternating current generator or the secondary of a transformer, current will pass during the half of the cycle when the voltage is directed so that P is positive. However, when P is negative the electrons from F are repelled from it and pass back into the filament. No appreciable number of electrons can be emitted from the cold plate P, so no current can flow through the circuit during that half of the cycle. Thus current flows through the circuit only half of the time and is zero during the rest of the time. The circuit used in this manner is called a half-wave rectifier and the diode is then called a rectifier tube.

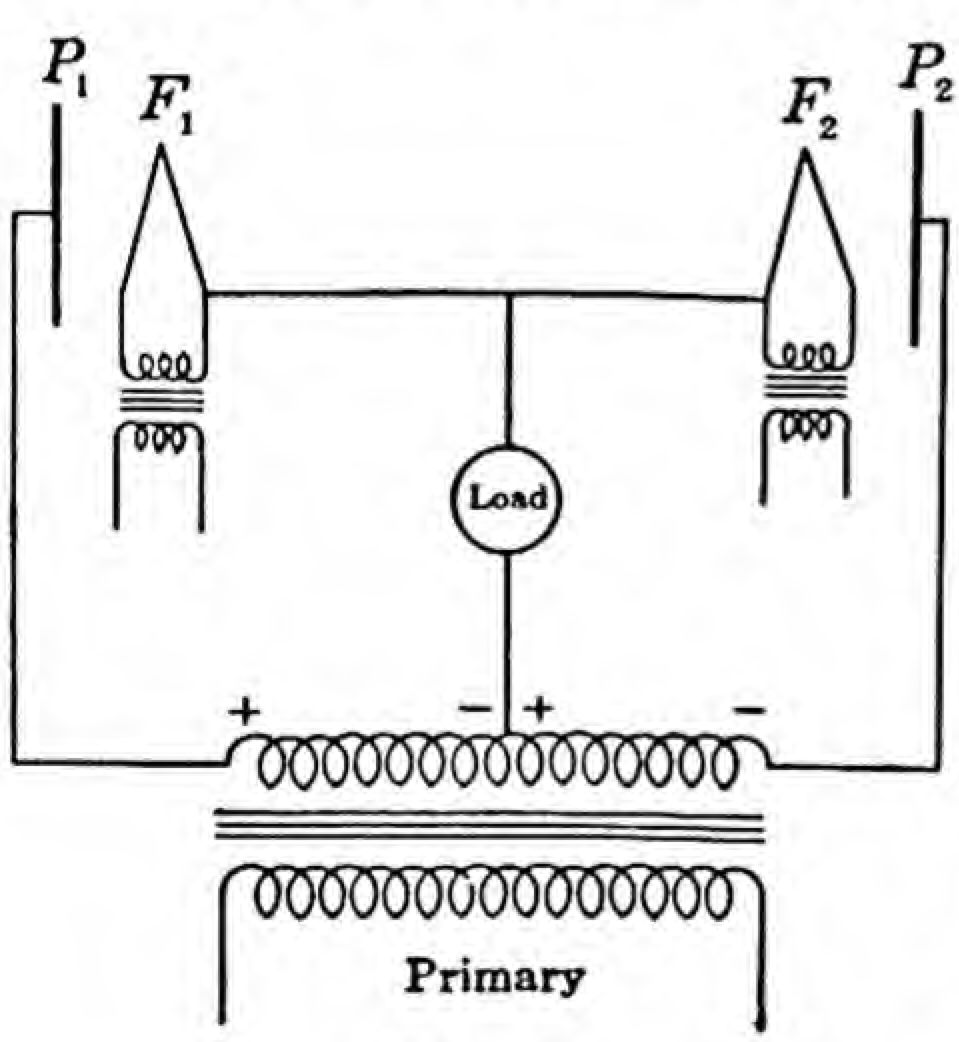


Fig. 225

Full-wave rectification may be obtained by using two tubes as shown in Fig. 225. The B voltage must be mid-tapped and is usually the secondary of a transformer. The filaments may be heated from low-voltage transformers. During the half of the cycle when the surge in the secondary of the main transformer is

as indicated in the figure, the plate P_2 is negative so no electrons can flow to it. However, the plate P_1 is positive with respect to its filament, so electrons flow from the mid-tap of the transformer up through the load and through the left-hand rectifier tube. During the other half of the cycle all potentials are reversed. The plate P_1 is negative with respect to F_1 so no current can flow in the left-hand circuit, but the plate P_2 is positive with respect to F_2 so current flows up through the load to F_2 and to P_2. Thus it is seen that first one half of the secondary of the transformer and then the other half supplies current through the load in the same direction. This load current although unidirectional is pulsating as shown in Fig. 226, the shape of each pulse depending on the wave form of the secondary of the transformer. Various combinations of condensers and choke coils placed in series with the load smooth out the wave, making a more steady direct current.

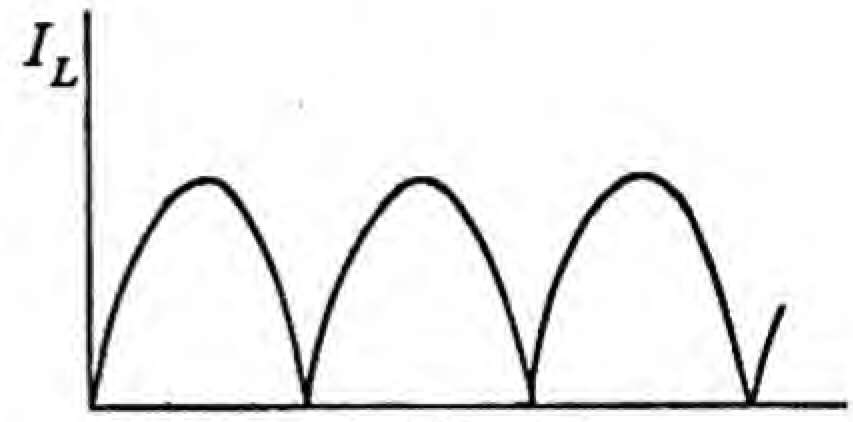


FIG. 226

245. The Triode. — We have seen that the space charge has a strong controlling action on the current through a diode and we can foresee that any means of varying it will be a means of varying the plate current. Let a wire or a grid of wires G be placed between the filament and the plate (Fig. 227) and let it be maintained by the battery C at a certain potential with respect to the filament. If the C battery is directed as shown, the grid is made negative with respect to the filament. This increases the space charge and hence decreases the plate current. If the C battery is reversed, the grid becomes positive with respect to the filament. This speeds up the electrons, decreases the space charge, and hence increases the plate current. Most of the electrons moving toward the grid pass through its open spaces, but those that hit the grid pass on through the C battery and form the very small grid current. The space charge being decreased, an increased current flows to the plate. Thus, by controlling the potential of the grid E_g, the current in the plate circuit is controlled. The

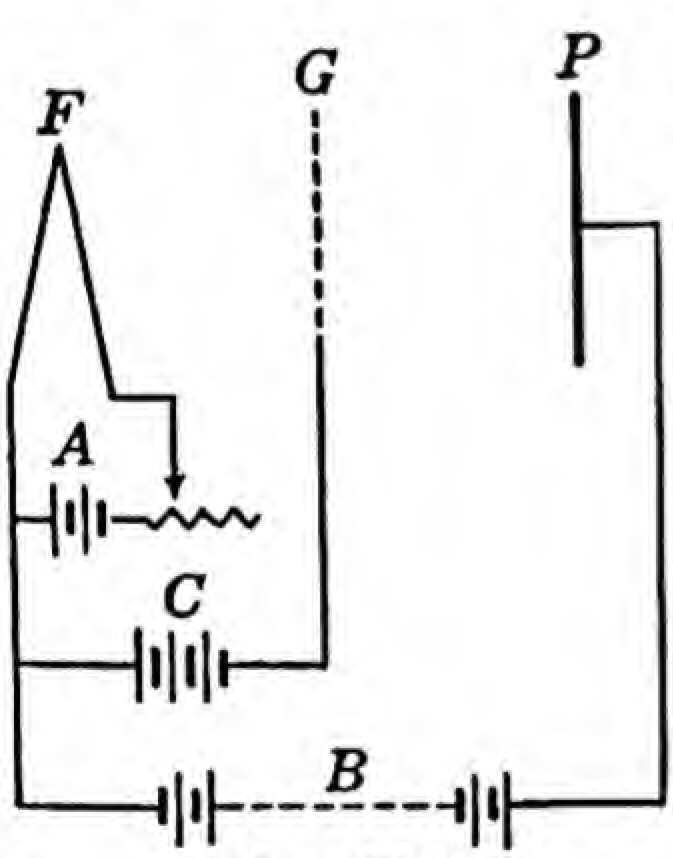


FIG. 227

relation between I_p and E_g (when the filament temperature is kept constant) is shown in Fig. 228. For a low value of E_p and zero grid potential the plate current is indicated by Ob. As E_g is made positive the current increases until the limit of emission of electrons from the filament is reached. With pure metal filaments the saturation values may be actually reached, but in the modern vacuum tube the oxide coated filaments emit electrons so copiously that the tube has to be badly overloaded before the bend in the upper part of the curves is reached. Upon reversal of E_g a certain negative value of E_g is reached at which and beyond which I_g is zero. As the plate potentials are increased it is seen that the values of E_p must be more and more negative in order to obtain the corresponding values of I_p. It is to be noted that for a considerable region each curve has a nearly straight portion.

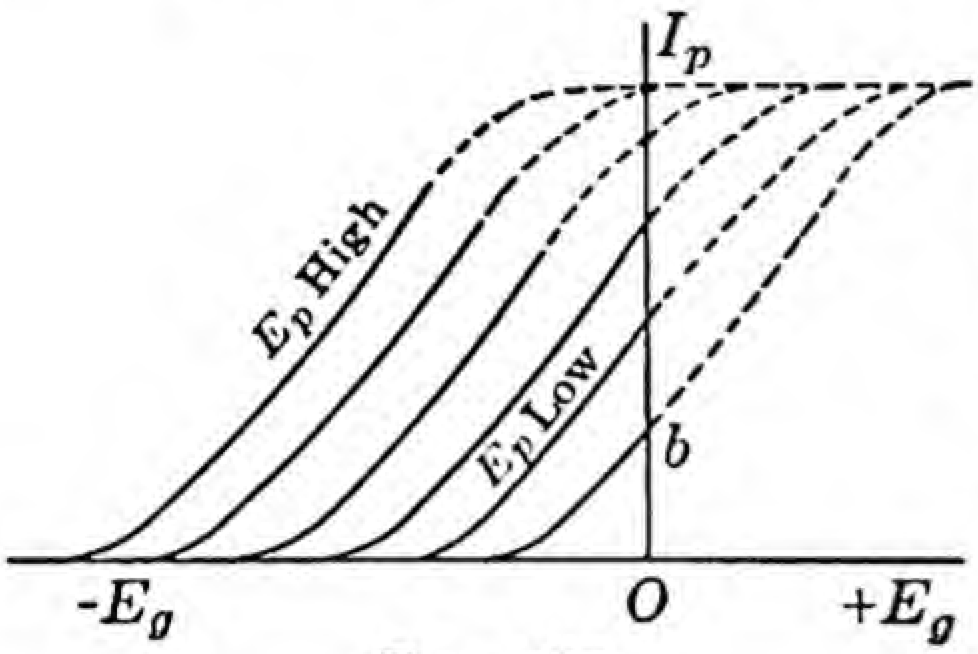


FIG. 228

246. The Triode as an Amplifier. — Suppose that a weak signal is received from a line as shown in Fig. 229 (for simplicity the filament batteries are not shown in the figure). The signal produces small changes in the potential of the grid, which in turn cause relatively large changes in the plate current. This fluctuating plate current causes a fluctuating voltage across the terminals of the

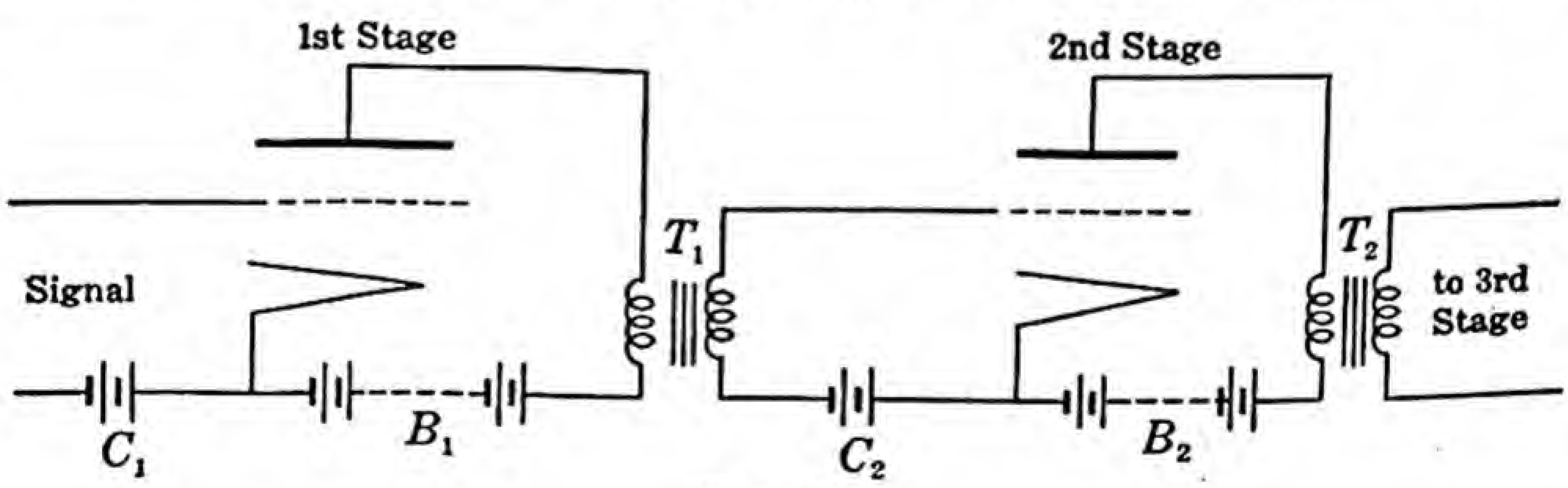


FIG. 229

transformer T_1 which is several times larger than the voltage of the incoming signal. In order to have the changes in I_p directly proportional to the strength of the incoming signal, the voltage of the C battery must be such as to keep the tube working on the straight-line portion of the operating curve. The transformer, in turn, steps up the voltage in direct proportion of the number of

turns on the secondary to the number of turns on the primary. The secondary of T_1 is connected across the grid and filament of the second vacuum tube where the process is repeated. As long as increased power is not desired, vacuum tubes are used which are designed to have small plate currents but have large voltage amplification. When large power amplification is desired, then in the final stage of the amplifier a tube is used which has a large plate current and which is designed for a high-voltage B battery.

It is to be noted that a single tube may increase the power enormously. As we have seen in the previous paragraph, the grid current is practically zero and the grid voltage is rather small so that the power required is small. However, a power tube may be designed so that the plate voltage is high and so that the small changes in grid voltage produce large fluctuations in the plate current, thus producing a large power output. The increased energy comes from the B battery, the grid acting only as a valve controlling the rate of flow of the energy.

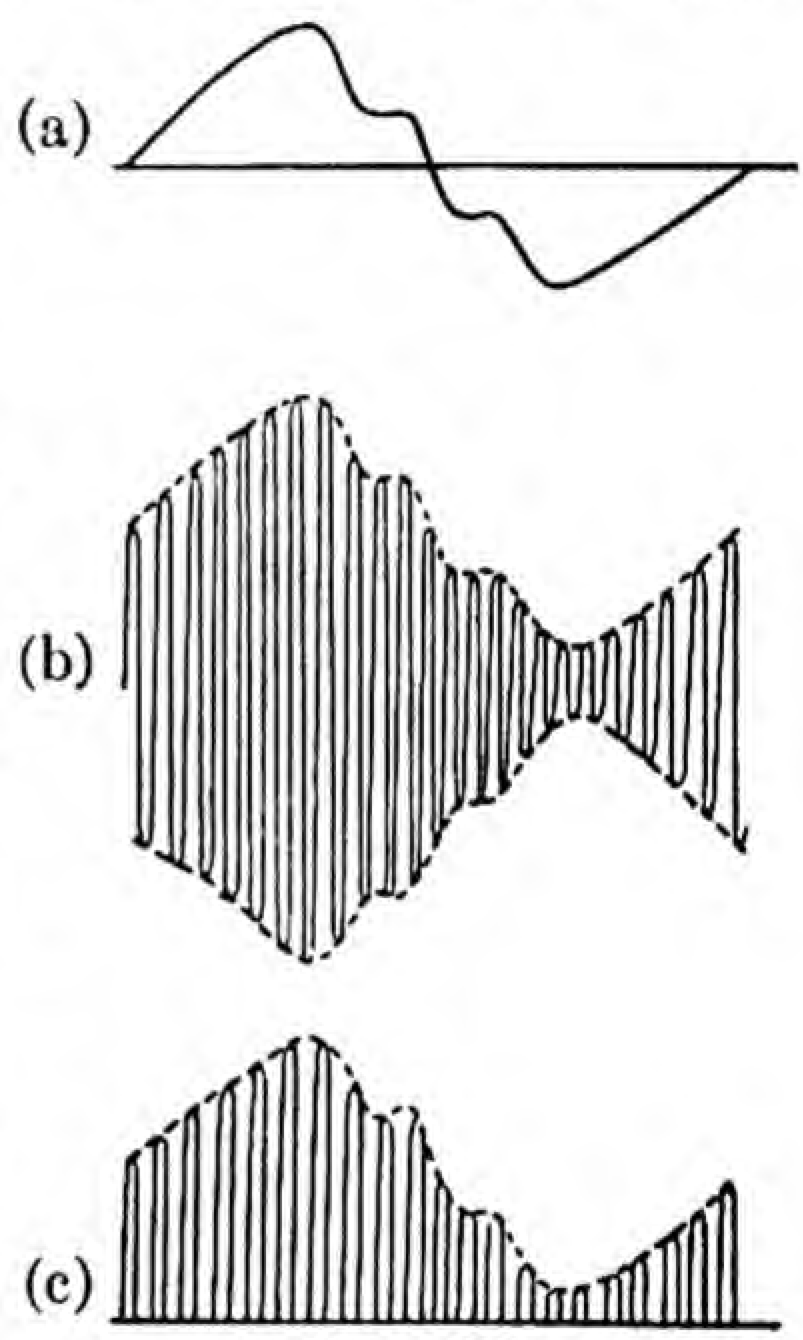


FIG. 230

247. Modulation.—In order to have a series of electromagnetic waves transmitted over any considerable distance, the frequency of the waves must be exceedingly high. We wish to explain how audible sounds may be transmitted by means of these high frequency waves. Consider a portion of a sound wave represented in Fig. 230*a* which it is desired to transmit. The broadcasting antenna when no audible signal is being sent is radiating waves of constant frequency and constant amplitude. In order to transmit the signal it modifies the amplitude of the waves as shown in Fig. 230*b* so that the envelope of the waves is exactly the shape of the signal to be sent. Now some device must be found which will pass the electric pulses in one direction and not pass them in the other direction, that is, transmit only the upper half of the waves *b*. Any such device is called a detector. The rectified wave for the above case is indicated in Fig. 230*c*. Any receiver such as a telephone or a loud-speaker which had such a series of

half-wave pulses passing through it would create a sound like the original. The separate impulses are of such very high frequency that the diaphragm could not follow them but would move only according to the average of the impulses, and thus produce the original wave motion. Obviously the frequency of the "carrier wave" must be ever so many times greater than the highest audible frequency to be transmitted.

248. Detectors. — It will be seen at once that a detector is nothing more than a rectifier and hence the diode may be used as a detector. The wires carrying the incoming high frequency signal, amplified or not, are connected into the circuit shown in Fig. 222 in place of the B battery. A pair of headphones or an audio-frequency amplifier replaces the milliammeter.

The triode may also be used as a detector. It is connected in the circuit just as the first amplifier tube in Fig. 229. However, the grid is made more negative with respect to the filament than when the tube is to be used as an amplifier. As shown in Fig. 231, the grid potential is made equal to OA, the cut-off point. When the incoming signal makes the grid less negative (point B), the value of I_p is nearly proportional to the magnitude of the signal voltage. When the signal voltage reverses, the grid becomes more negative and no current flows. In this respect the tube is a rectifier much as the diode. However, there is amplification as well as rectification in the triode. Thus in Fig. 229, if the signal is such as shown in Fig. 230b, the first tube passes on to the transformer T_1 a wave such as the envelope in Fig. 230c, rectified and amplified. The iron transformer chokes out the high-frequency changes so that the current through the rectifier circuit is of the form in Fig. 230a.

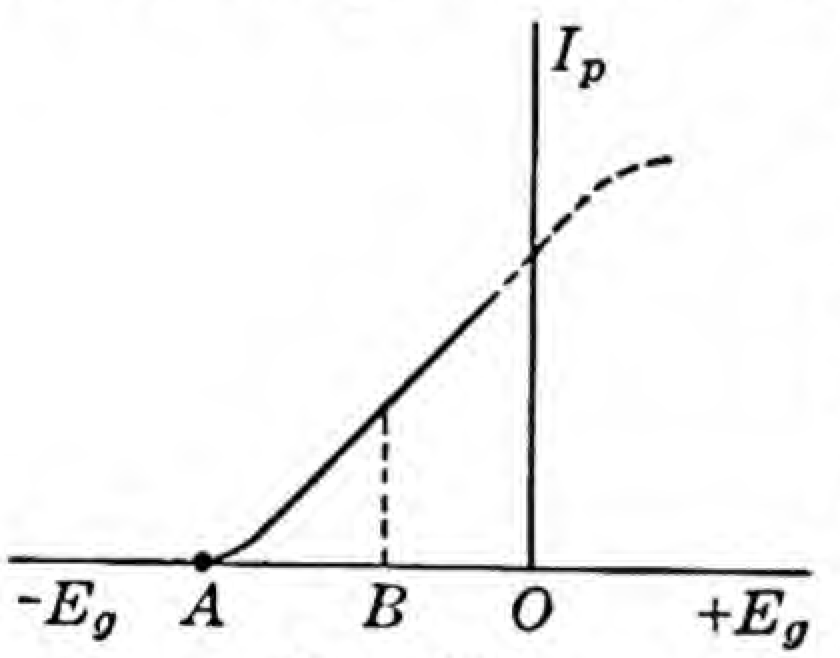


FIG. 231

249. The Vacuum Tube Oscillator. — In order to broadcast audible sounds we have seen that we must modulate a series of very high-frequency waves of constant amplitude. To produce these waves an oscillator must be made to supply high-frequency currents. The triode is well adapted for this use. Means must be provided for producing the oscillations in the plate circuit and there must be a feed-back from the plate circuit into the grid cir-

cuit so that the change in grid potential tends to boost the oscillations.

The Hartley circuit shown in Fig. 232 offers one of the simpler types of oscillators. L_1 and L_2 are inductances so placed that there is flux linkage between them. Any slight change in the plate current causes a change in flux in L_1 and hence in L_2. The change in flux in L_2 causes an induced e.m.f. which produces a change in the grid bias of the vacuum tube and, in turn, causes a further change in the plate current. Now the lower portion of the circuit is a parallel resonant circuit in which electrical oscillations will occur with a period given approximately by $T = 2\pi\sqrt{LC}$ where L is the total self-inductance of L_1 and L_2. Every oscillation in the resonant circuit is boosted by the action of L_2 upon the grid. So a continuous set of oscillations of constant amplitude is maintained in the plate circuit.

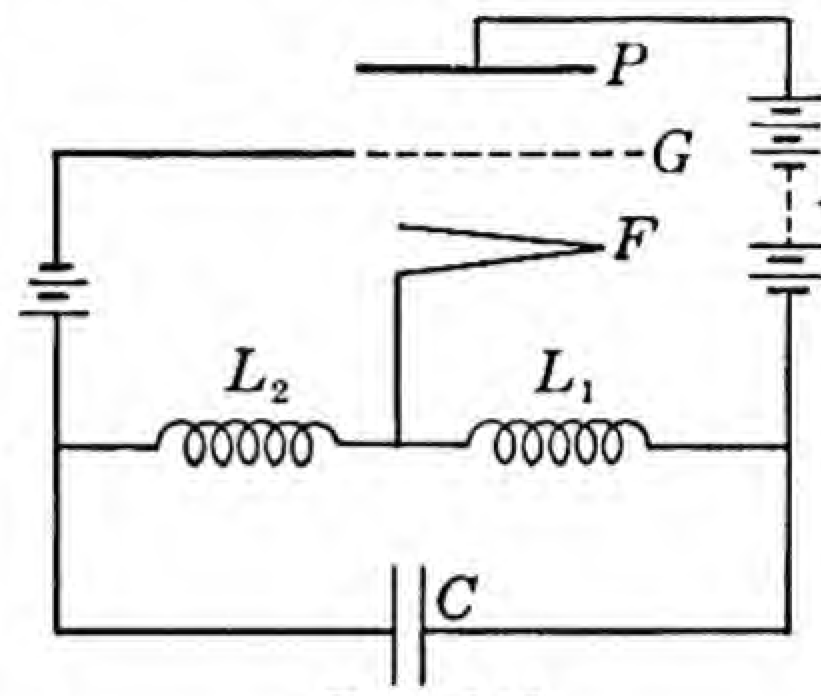


FIG. 232

THE QUANTUM THEORY

250. The Energy Quantum. — At the close of the 19th century there were a number of outstanding problems in physics which the existing theory (now called the classical theory) was utterly unable to explain. The spectral distribution of energy in the radiation from a black-body (or uniform temperature enclosure), the relation of the specific heat of a body to its temperature, the photoelectric effect, and the origin of spectra offered insuperable difficulties. These are primarily problems in the radiation, absorption, and transfer of energy.

Maxwell's electromagnetic theory (published in 1864) had placed the discoveries of Faraday on a sound mathematical basis. His theory predicted the existence of electromagnetic waves and suggested that light must consist of such waves. The various properties of light, interference, polarization, dispersion, etc., were satisfactorily explained and the exact velocity of light predicted. Electromagnetic waves from oscillating electric charges were experimentally produced by Hertz in 1887, verifying Maxwell's prediction and thoroughly establishing the theory.

The failure of the classical theory to explain the phenomena above-mentioned led to more searching inquiry. In the year 1900 Max Planck suggested that the trouble might be found in our view of the nature of radiation itself. He suggested that the atomic oscillators (electrons or atoms) possibly did not radiate in a continuous manner as required by Maxwell's theory but released energy only in certain units of energy (quanta), in other words that energy was atomic in its nature. Further, these energy units were considered to be proportional to the frequency ν of the radiation. Thus

$$W_r = h\nu. \tag{336}$$

h is known as Planck's constant.

This revolutionary assumption, disturbing as it was in its contradiction of the wave theory, enabled Planck to derive the correct radiation law, led Einstein to a clear and simple explanation of the photo-electric effect, Einstein and Debye together to the correct specific heat equation, and Niels Bohr to an explanation of the origin of spectra. While these and many other difficulties have been explained by this concept, the quantum theory has in itself introduced a most serious problem in its apparent contradiction to the wave theory of light. The complete reconciliation of these two theories is still to come.

251. The Radiation Law. — The nearest approach to the radiation law afforded by classical physics is found in the expression

$$E_\lambda = C_1\lambda^{-5} e^{-\frac{C_2}{\lambda T}}. \tag{337}$$

This is known as Wien's law (see § 112). The heavy line (Fig. 233) represents the experimental data. The dotted line shows how near Wien's law fits the data. The discrepancy at long wave-lengths indicated that something was wrong with the theory.

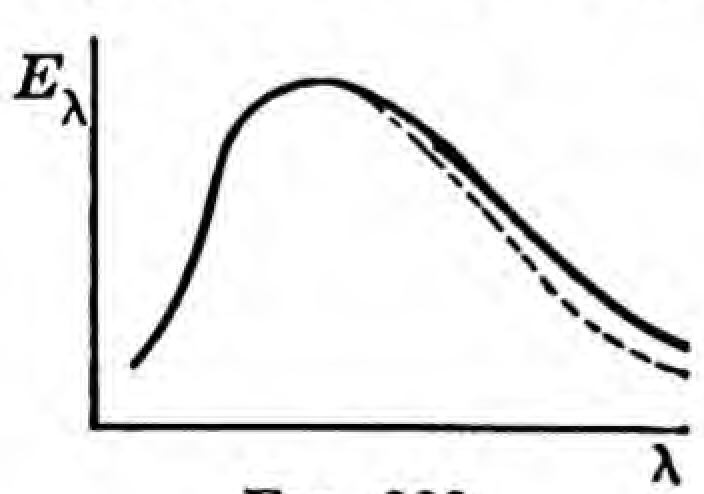


Fig. 233

It is well established that all radiation (light or radiant heat) is electromagnetic in nature, — that radiation consists of electromagnetic waves. In producing electromagnetic waves of the broadcasting range we set up systems of oscillating electric charges and these oscillating charges radiate the waves which our receivers pick up. Just so, for light waves we think of oscillators of atomic dimensions, the charged particles

(ions and electrons) of which the atoms are composed. We think of the motions of these oscillators as increased under thermal agitation, that is, with rise of temperature. Under the classical view these oscillators possess all gradations of energy from zero to infinity according to Maxwell's probability law (Fig. 234) which gives the distribution of energy among the molecules of a gas. These electrical oscillators must, therefore, radiate energy in all gradations. This view led to Wien's radiation law. Planck suggested that possibly energy cannot be possessed by these oscillators in all gradations but that it must be possessed in multiples of some fundamental unit which has to be associated with the frequency by the relation

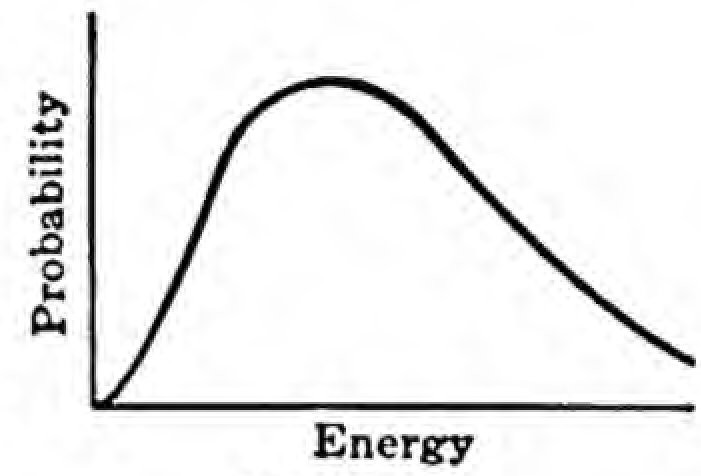


FIG. 234

$$W = h\nu \tag{338}$$

where ν (nu) is a frequency; h, a constant.

This idea calls for energy units (quanta) of different sizes, corresponding to the frequency. Planck suggested that not only is energy possessed in discrete quanta by the oscillators but that it must be radiated in discrete quanta. This view leads to a radiation law which exactly fits the experimental facts. The law, whose development is beyond the scope of this book, takes the form

$$E_\lambda = C\lambda^{-5}\frac{1}{e^{\frac{h\nu}{kT}} - 1}. \tag{339}$$

In 1905 Einstein offered an explanation of the photo-electric effect on the assumption that radiation is absorbed in quanta, thus extending Planck's idea which suggested its emission in quanta. The photo-electric effect has been discussed in § 188.

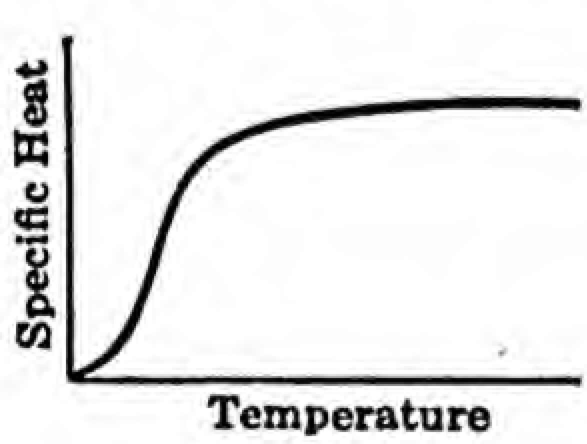


FIG. 235

The next application of the quantum idea was also made by Einstein (1907) and directed toward the problem of the change of specific heat with temperature. The change of specific heat with temperature as found experimentally is shown in Fig. 235. On the basis that energy may be transferred only in discrete quanta, Einstein developed an equation in fairly close agreement with the facts. Certain modifications by Debye (1912) brought the equation into still better agreement.

252. The Structure of the Atom. — Sir Ernest Rutherford in 1911 offered experimental evidence in favor of the so-called nuclear atom, that is, an atom composed of electrons revolving about a minute central nucleus. The idea was expanded by Niels Bohr in a very celebrated paper in the Philosophical Magazine in 1913. In this paper was developed a complete description of the hydrogen atom and the manner in which its spectrum was produced. In order to understand Bohr's theory it will be necessary to consider first certain experimental facts concerning spectra, particularly the hydrogen spectra.

253. Spectral Series. — The spectrum of any element, even that of hydrogen, consists of very many lines extending over the ultraviolet, the visible, and the infra-red regions. That many of the lines formed series indicating relationships which could be expressed by formulae was early recognized. The earliest of these series relationships was one discovered by Balmer in 1885, viz.,

$$\lambda = k \frac{m^2}{m^2 - a^2}. \tag{340}$$

In this formula $a = 2$, k is a constant, and m has successive values 3, 4, 5, . . . Rydberg (1890) developed a more general formula applicable to all series (of a certain type) of which Balmer's formula was a special case. In terms of wave-numbers (number of waves per centimeter), Rydberg's formula is

$$n_m = n_\infty - \frac{R_\infty}{(m + \mu)^2}. \tag{341}$$

R_∞ is a universal constant (the Rydberg constant); n_∞ and μ are constants, characteristic of a given element; m is the ordinal number of a given spectral line. As m increases the formula requires the lines to crowd closer and closer together approaching n_∞ as a limiting value. n_∞ is thus the highest or convergence wave-number and corresponds to the shortest wave-length in the series. There are several series of the Rydberg type for a given element characterized by different convergence wave-numbers n_∞ and different values of μ. It is found that the convergence wave number for one series is the first or second term of some other series so that the formula may be written

$$n_m = \frac{R_\infty}{(1 + \mu')^2} - \frac{R_\infty}{(m + \mu)^2}. \tag{342}$$

If we multiply each term in this equation by ch, where c is the velocity of light and h is Planck's constant, the terms become energy quantities. Thus since $cn = \nu$, where ν is the frequency, we have

$$h\nu = \frac{chR_\infty}{(1 + \mu')^2} - \frac{chR_\infty}{(m + \mu)^2}. \qquad (343)$$

This equation is highly significant. The left-hand member is the energy quantum radiated corresponding to frequency ν. The right-hand members must be energy states and hence the quantum radiated is the difference between two energy states within the atom. This is the fundamental relation upon which any theory of atomic structure must be based. Eq. 343 may be written more simply as

$$h\nu = W_1 - W_2. \qquad (344)$$

254. The Bohr Atom. — The hydrogen atom of Bohr is the Rutherford atom consisting of a central nucleus and an electron revolving in circular (or elliptical) orbits about it. For some reason (unknown) only certain orbits are allowable. An electron in a given orbit represents a particular energy state. The normal state is, in accordance with the laws of mechanics, the state of lowest potential energy and this is the allowable orbit nearest the nucleus. When an atom absorbs energy, as it may by a collision with another atom or with an electron or from radiation incident upon it, the electron is thrown to a higher energy orbit, farther from the nucleus. By whatever means it gets energy it may absorb it only in quanta, $h\nu$, the amount necessary to put the atom in some other allowable energy state. On the Bohr view this is the energy required to throw the electron either into one of the allowable outer orbits or clear out of the atom. If the atom is thus put into a higher energy state, it is said to be *excited*. If the electron is thrown clear out of the atom, the atom is left with a positive charge and is said to be *ionized*. In either condition the atom will tend to recover its equilibrium, that is, to return to its lowest energy state, and it does this not by gradual release and radiation of energy but by abrupt jumps to lower states. Since, by Eq. 344, the successive energy states differ by discrete quanta, each of value $h\nu$, the radiation must be in discrete quanta, and the frequencies of the various spectral lines must be determined by the differences of the various allowable energy states.

Let us assume not only that energy is absorbed and released in quanta as called for by Eq. 344 but that the states themselves are "quantized," that is, consist of multiples of $h\nu'$, where ν' may be the frequency of the electron in its orbit (not to be confused with ν, the frequency radiated or absorbed). If we thus assume that

$$W = \tau h\nu', \tag{345}$$

where τ is any whole number 0, 1, 2, 3, . . ., we can then determine the allowable orbits as follows. We may write $\tau h\nu' = I\omega^2$, where I is the moment of inertia of the electron and ω ($= 2\pi\nu'$) the angular velocity of the electron in its orbit. The total energy of the system which is thus quantized is twice the kinetic energy, $\frac{1}{2} I\omega^2$, since there is always an equal amount of potential energy.* Now the momentum of the electron is $p = I\omega$. Hence $I\omega^2 = p\omega = \tau h\nu'$ or $p = \tau h/2\pi$, since $\omega = 2\pi\nu'$. The condition $p = \tau h/2\pi$ enables us to compute the radius of an orbit designated by any particular quantum number τ. Thus angular momentum $p = I\omega = mr^2\omega$, so that

$$mr^2\omega = \frac{\tau h}{2\pi}. \tag{346}$$

Assume the central force acting on the electron is given by the inverse square law. Then

$$\frac{Ee}{r^2} = m\omega^2 r,$$

where E is the charge on the nucleus, e the charge on the electron. This may be written

$$\frac{Ee}{2r} = \frac{1}{2} mr^2\omega^2.$$

This is an expression for kinetic energy. Eliminating ω by use of Eq. 346 and solving for r, we have

$$r = \tau^2 \frac{h^2}{4\pi^2 mEe}.$$

Thus the radii of successive orbits are proportional to the squares of the numbers 0, 1, 2, 3, . . . etc.

* We are considering a case of uniform circular motion. Such motion is equivalent to two superposed linear simple harmonic motions of equal amplitudes and periods, at right angles and 90° out of phase. When the energy associated with one component is all kinetic, there is an equal amount of energy, which is all potential, associated with the other component. The total energy is thus $I\omega^2$.

The total energy of the system in any state is $W = K.E. + P.E.$ The kinetic energy $K.E. = Ee/2\,r$ as shown above. The potential energy, defined as the work required to bring the negative charge $-e$ from infinity, is $P.E. = -eE/r$ (§ 119). The total energy thus becomes

$$W = -\frac{Ee}{2\,r}.$$

Thus a given energy state is expressed by

$$W = -\frac{1}{\tau^2}\frac{2\,\pi^2 mE^2e^2}{h^2}.$$

Our equation, $h\nu = W_1 - W_2$, becomes

$$h\nu = \frac{2\,\pi^2 mE^2e^2}{h^2}\left(\frac{1}{\tau_1{}^2} - \frac{1}{\tau_2{}^2}\right).$$

Expressed in wave numbers, by dividing by hc, this relation becomes

$$n = \frac{2\,\pi^2 mE^2e^2}{ch^3}\left(\frac{1}{\tau_1{}^2} - \frac{1}{\tau_2{}^2}\right). \qquad (347)$$

This is the famous Bohr equation for the hydrogen atom.

Let us compare this with the Balmer equation which may be written in the form

$$n = R_\infty\left(\frac{1}{a^2} - \frac{1}{m^2}\right),$$

in which $a = 2$ and $m = 3, 4, 5, \ldots.$ Thus the Rydberg constant for the hydrogen atom is

$$R_\infty = \frac{2\,\pi^2 me^4}{ch^3},$$

since $E = e$ (for hydrogen).

From the known value of the constants on the right-hand side

$$R_\infty = 109{,}750.$$

The value obtained experimentally from data on spectra is

$$R_\infty = 109{,}678.$$

This close agreement is one of the greatest triumphs of the quantum theory and lends great plausibility to the Bohr atom, which after all is a very artificial conception. When we consider elliptical orbits, the effect of the nucleus, and correct the mass (m) of the electron in a way required by relativity theory for high

velocities, the agreement is closer and even the detailed structure of the spectral lines is accounted for.

One of the difficulties with the theory is the assumption that the revolving electrons do not radiate, the radiation occurring only during transitions from one orbit to another. In spite of the remarkable success of the Bohr theory we are today adopting a more cautious view and prefer to speak of energy states retaining the quantum transitions but putting a less literal interpretation on electron orbits.

INDEX

References are to page numbers.